Theoretical and Practical Aspects of Strongly Coupled Aerosol-Atmosphere Data Assimilation

Ting-Chi Wu, Milija Zupanski, Anton Kliewer, Lewis Grasso, and Leah D. Grant

Abstract Coupled data assimilation is one of the most active research areas in recent years because of its potential for improving the prediction of coupled modeling systems. Among various coupling options, strongly coupled data assimilation is the most efficient option for processing the information from observations. At the same time, coupled aerosol-atmosphere modeling is steadily gaining more interest due to its relevance to air quality, aviation, solar energy, and climate. It is well known that aerosols play an important role in Earth's radiation balance. Aerosol-atmosphere interaction is clearly multi-scale, from large-scale stratospheric impact to small-scale aerosol-cloud interaction. Such complex prediction system requires advanced data assimilation methodology that can deal with multi-scale interactions and observation information flow. In this chapter we address theoretical and practical aspects of strongly coupled data assimilation in application to aerosol-atmosphere coupling. We describe major aspects of developing strongly coupled data assimilation and related challenges. We also show results from a case study using a recently developed regional aerosol-atmosphere coupled data assimilation system. Finally, a general discussion on the future needs of strongly coupled data assimilation is provided.

T.-C. Wu (✉) · M. Zupanski · A. Kliewer · L. Grasso
Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO 80523, USA
e-mail: ting-chi.wu@colostate.edu

A. Kliewer
NOAA Earth System Research Laboratory, Global System Laboratory, Boulder, CO, USA

L. D. Grant
Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA

1 Introduction

The primary role of coupling of prediction systems is to allow more realistic interactions between previously independent components, and therefore have a more accurate representation of relevant dynamical and physical processes. Since data assimilation has been an integral part of numerical weather prediction (NWP), there is a need for developing data assimilation for coupled prediction systems, often referred to as *coupled data assimilation.* A commonly used classification of coupled data assimilation includes *weakly* and *strongly* coupled data assimilation (Penny et al. 2017; Zupanski 2017). In a *weakly* coupled system, each component (e.g., atmosphere, chemistry, aerosol) has its own independent data assimilation system and analysis. In a *strongly* coupled system, all coupled system components are included in a holistic data assimilation system that can simultaneously assimilate observations from all components.

1.1 Background on Coupled Data Assimilation System

In this chapter we are primarily interested in describing strongly coupled data assimilation in aerosol-atmosphere coupled prediction system. Commonly used coupled aerosol-atmosphere prediction systems include the Goddard Earth Observing System Version 5 (GEOS-5), Navy Global Environmental Model/Navy Aerosol Analysis and Prediction System (NAVGEM/NAAPS), European Centre for Medium-Range Weather Forecasts/Copernicus Atmosphere Monitoring Service (ECMWF/CAMS), National Oceanic and Atmospheric Administration (NOAA) Global Forecast System (GFS), the Weather Research and Forecasting-Chemistry (WRF-Chem), and the Regional Atmospheric *Modeling* System (*RAMS*) *Model* (Molod et al. 2012; Hogan et al. 2014; Morcrette et al. 2009; Putman and Lin 2007; Chen et al. 2013; Grell et al. 2005; Fast et al. 2006; Saleeby and van den Heever 2013). Although some aspects presented here may be of general importance for data assimilation, they are mainly relevant to commonly used variational, ensemble, and hybrid variational-ensemble data assimilation systems (Parrish and Derber 1992; Rabier et al. 1999; Houtekamer and Mitchell 2001; Whitaker and Hamill 2002; Kleist and Ide 2015). In those systems, the background (or sometimes referred to as forecast or prior) error covariance is a key element of successful data assimilation analysis (e.g., Lorenc 1986; Kalnay 2003), which directly implies that coupled background error covariance plays a fundamental role in coupled data assimilation. Further, cross-covariance between components in a coupled system has the same relevance as the cross-covariance between variables in a standalone system. For example, it is well known that there exists a physical relationship between atmospheric temperature and wind. Data assimilation that includes such correlations (or cross-covariance) between wind and temperature in its background error covariance will produce more accurate analysis than a standalone data assimilation for wind and for temperature. Similarly, if correlations between an aerosol

and an atmospheric variable exist, a data assimilation that includes such correlations in the coupled background error covariance will produce more accurate analysis.

Another benefit of strongly coupled data assimilation is that it provides a mechanism for transferring observation information between coupled components. This may be especially relevant for coupled aerosol-atmosphere system. Given that there are generally fewer aerosol observations than atmospheric observations, assimilation of atmospheric observations can potentially improve aerosol initial conditions. Atmospheric observations can also be beneficial for improving the vertical distribution of the aerosol initial conditions, even when aerosol observations are assimilated. The most widely available aerosol observations are in terms of Aerosol Optical Depth (AOD), which are a vertically integrated quantity and therefore do not produce a vertical distribution of aerosol. In that situation, using observed atmospheric profiles can provide additional information about vertical distribution of aerosol through strongly coupled data assimilation.

1.2 Theoretical Description of Coupled Data Assimilation System

In order to illustrate the impact of coupled data assimilation, we consider a two-variable, one-point, aerosol-atmosphere coupled system. As shown in Zupanski (2017), when atmospheric component is observed under such system, the Kalman filter analysis equation can be written as follows:

$$x_{atm}^a = x_{atm}^b + \frac{\varepsilon^2}{1+\varepsilon^2}\left[y_{atm} - x_{atm}^b\right] \tag{1}$$

$$x_{aero}^a = x_{aero}^b + \rho\left(\frac{\sigma_{aero}}{\sigma_{atm}}\right)\frac{\varepsilon^2}{1+\varepsilon^2}\left[y_{atm} - x_{atm}^b\right] \tag{2}$$

$$\varepsilon = \frac{\sigma_{atm}}{r_{atm}} \tag{3}$$

In the above equations, subscripts *atm* and *aero* refer to atmospheric and aerosol components, respectively, superscripts *a* and *b* denote analysis and background, respectively, x is state, y is observation, σ and r denote background and observation errors, respectively, and ρ is the correlation between atmospheric and aerosol variables. Equation (1) is a standalone analysis for the atmospheric component, which means that when only atmospheric variables are observed, the coupled atmospheric analysis is identical to the standalone atmospheric analysis. Equation (2) represents the aerosol analysis, which critically depends on the correlation between atmospheric and aerosol variables (Eq. 3). When the correlations between atmospheric and aerosol variables are non-existent or negligible, aerosol analysis is the same as the guess,

meaning no change from the assimilation. However, when the correlations exist the aerosol analysis can be updated from assimilating atmospheric observations.

The above discussion illustrates the main motivation for using the formalism of strongly coupled data assimilation instead of weakly coupled data assimilation: strongly coupled data assimilation is more general as it includes weakly coupled assimilation as an option. When correlations between variables are naturally negligible, a strongly coupled system will still correctly produce the analyses approximately equal to standalone analyses. When correlations are relevant, the strongly coupled system will update all variables, effectively increasing the utility of observations. The implied assumption for achieving the desired impact of strongly coupled data assimilation is that the estimated cross-correlations are reliable.

One critical issue in strongly coupled data assimilation is related to spatial and temporal scales of coupled processes. Although further understanding of the impact of having different spatial and temporal scales between a coupled system on the estimate of the background error covariance is necessary, it is likely that in an idealized data assimilation scenario where error covariances are exact and full-rank, all correlations (temporal, spatial, cross-variable, and cross-components) will be accurately accounted for. This is because in that situation the covariance would accurately represent the interactions between uncertainties of coupled components, and therefore implicitly address the scale differences. In practical applications, however, the coupled error covariance may not be able to account for different scales of coupled components (e.g., aerosol and atmosphere) with sufficient accuracy, in particular the temporal scales. While there is no commonly accepted solution to this problem, a possible strategy in such situations could be to modify existing background error covariance to reflect the different temporal scales between coupled components. For example, one could enforce covariance localization in time using pre-defined characteristic correlation scales or one could also use a covariance averaged over several previous data assimilation cycles. That said, the aerosol and atmosphere time scales may not be as different as the time scale differences between other coupled system such as the land surface and the atmosphere. As such, accounting for temporal correlations may not be a concern in an aerosol-atmosphere coupled system. In any case, incorporating different time scales in coupled error covariance is an important next step in making strongly coupled aerosol-atmosphere data assimilation more reliable and effective.

When using variational data assimilation, in which error covariance is approximated by a mathematical function, satisfactory modeling the correlations between coupled components may be difficult to achieve (Ménard et al. 2019). However, aerosol and chemistry data assimilation with four-dimensional variational (4D-Var) methods may offer new possibilities. For example, Hakami et al. (2005) found that adjoint inverse modeling in 4D-Var helps in constraining various inputs for chemical transport models, while Sandu et al. (2005) concluded that 4D-Var is a feasible approach for carbon-cycle aerosol assimilation. As a smoother, 4D-Var has the advantage of automatically accounting for time correlations during the data assimilation process. On the contrary, time correlations have to be fully imposed in sequential data assimilation, i.e. filters. With that, 4D-Var can be an advantageous option

for coupled aerosol-atmosphere data assimilation since the interaction between the different time scales of aerosol and atmosphere will be more realistic in 4D-Var compared to the interactions in filters. This certainly opens additional avenues for strongly coupled aerosol-atmosphere data assimilation research directed towards using smoothers instead of filters.

When using ensemble data assimilation, however, all correlations come naturally from ensemble forecasting. A potential difficulty may be that small ensemble size does not produce reliable estimates of correlations, which then requires additional attention. Considering the above possibilities, it seems that using a strongly coupled data assimilation formalism has more advantages than disadvantages. Most importantly, strongly coupled formalism potentially allows a more efficient use of observations, eventually leading to an improved analysis and prediction.

1.3   Single Observation Experiment

One of the main advantages of using an ensemble data assimilation algorithm is the flow-dependent background error covariance. Created by ensemble of model forecasts it is time-dependent and includes complex correlations between variables. For aerosol data assimilation the correlations between atmospheric and aerosol variables have the most significance. In principle, the correlations allow observations of one component to impact the analysis of another component. This also helps in the areas where AOD observations may have insufficient coverage, by indirectly providing additional information through cross-correlation. Atmospheric observations also provide additional information about the three-dimensional structure of aerosol, through the flow-dependent correlations.

To illustrate this impact, we conduct two single observation experiments using a regional coupled chemistry-aerosol-atmosphere WRF-Chem model, with the Goddard Chemistry Aerosol Radiation and Transport (GOCART) aerosol module. The data assimilation interval is 6 h, and model grid spacing is 9 km with a total of 50 vertical layers.

In the first experiment, we assimilate a single east–west wind component (u wind) observation at 25°N, 53°E and near the model surface. In Fig. 1, we show the impact of such assimilation on the DUST_3 (2.4 µm) variable from the GOCART aerosol module. Note that in a less advanced, uncoupled data assimilation system, the impact of assimilating a single wind observation on dust variable would be equal to zero. In Fig. 1a, one can see negative increments of dust, in both horizontal and vertical directions, suggesting that increasing westerly wind in that area will produce a decrease of dust concentration. In Fig. 1b, one can also notice that the impact of wind observation on dust is limited in the vertical direction and is generally confined to lower levels where the observation was located.

In the second experiment, we assimilate a single DUST_3 observation in the same place located at 25°N, 35°E. The impact of assimilating such observation to near-surface wind is shown in Fig. 2. One can notice a dominant negative response, which
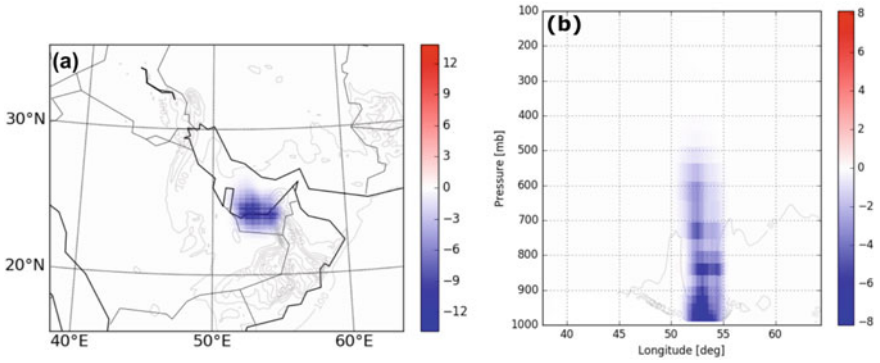
Fig. 1 Analysis increments (i.e., analysis minus background) of DUST_3 (ug kg^{-1}$_{dry\ air}$) in response to a single east–west wind observation (u component wind), valid at 00 UTC on August 4, 2016: **a** horizontal distribution at surface and **b** vertical cross section along 25°N
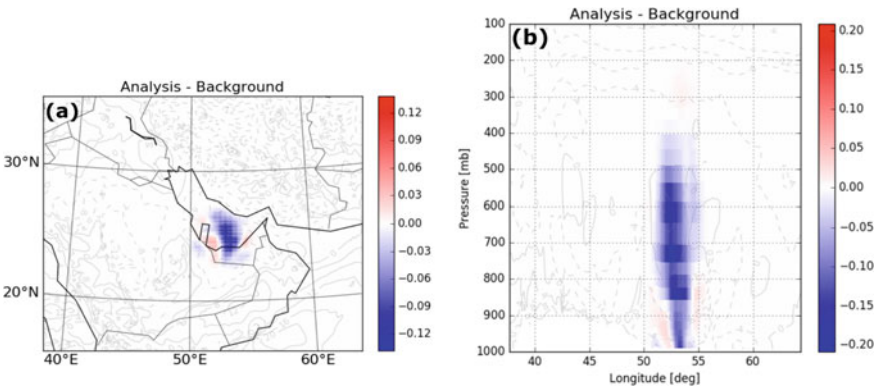

Fig. 2 Similar to Fig. 1, except for analysis increments of u component wind (m s^{-1}) in response to a single DUST_3 observation, valid at 00 UTC on August 4, 2016: **a** horizontal distribution near surface and **b** vertical cross section along 25°N

is consistent with the findings in Fig. 1. The analysis response of dust is also limited in both vertical and horizontal directions, as anticipated due to the use of covariance localization.

The rest of the chapter is organized as follows. We begin by describing the current status of aerosol-atmosphere coupled data assimilation in Sect. 2, followed by aerosol observations and observation operator in Sect. 3. Challenges of strongly coupled data assimilation are discussed in Sect. 4, with numerical experiments of a case study and results presented in Sect. 5. Summary and future directions are given in Sect. 6.

2 Current Status on Aerosol-Atmosphere Coupled Data Assimilation

Before an overview on the current status of aerosol-atmosphere coupled data assimilation is given, a brief discussion on a prerequisite topic regarding online versus offline approaches for weather and aerosol *forecasting* is provided herein. An offline approach involves an aerosol model run that is driven by meteorological fields produced by an atmospheric model run (e.g., Sekiyama et al. 2010; Rubin et al. 2017). As a result, interactions between the atmospheric and the aerosol processes are restricted to one-way. That is, the meteorological fields from an atmospheric model are used to initialize the aerosol model, but the outcome from the aerosol model is not fed back to the atmospheric model. On the other hand, an online (sometimes also referred to as inline) approach involves an integrated model run of both atmospheric and aerosol components (e.g., Liu et al. 2011; Lee et al. 2017), in which a two-way interaction of atmospheric and aerosol components is allowed. As indicated in Grell and Baklanov (2011), major advantages of using an online approach as opposed to an offline approach include a more realistic presentation of the atmosphere, a more numerically consistent treatment of both components, and improved forecast via improved assimilation. Nevertheless, the reduced computational cost and more flexibility in ensemble forecasting makes the offline approach still rather appealing, especially for regulatory agencies.

As mentioned in the introduction, there exist two general approaches for aerosol-atmosphere coupling from the *data assimilation* perspective. As discussed earlier, the two approaches are (i) a weakly coupled data assimilation and (ii) a strongly coupled data assimilation. A weakly coupled data assimilation system performs data assimilation of each component independently, although the updated analysis of both meteorological and aerosol fields can be used to initialize a coupled aerosol-atmosphere forecast. Since the individual component is treated separately, there does not exist cross-component elements in the background error covariance matrix, which is essential for the data assimilation update. In contrast, a strongly coupled data assimilation system performs data assimilation and forecast of both aerosol and atmospheric components simultaneously, treating the coupled system as a single integrated system. As such, there exist cross-component elements in the background error covariance matrix, which allows observational information from one component to potentially influence the other component within a coupled data assimilation update. Based on the varying degrees of data assimilation update, weakly (strongly) coupled data assimilation can be further classified into quasi weakly (strongly) and weakly (strongly). Interested readers are redirected to Penny et al. (2017) and Penny and Hamill (2017) for more details.

2.1 Operational Centers and Research Community

With the increased computational power, many NWP centers have reconsidered the *online* approach over the more common *offline* approach for weather and aerosol *forecasting*. For example, the ECMWF Integrated Forecast System (IFS) (Morcrette et al. 2008), the Japan Meteorological Agency (JMA) Model of Aerosol Species in the Global Atmosphere (MASINGAR) (Tanaka and Chiba 2005), and the UK Met Office (UKMO) Unified Model (UM) (Collins et al. 2011). Nevertheless, several NWP centers favor the offline approach and that include the US Navy Fleet Numerical Meteorology and Oceanography Center (FNMOC) Naval Research Laboratory (NRL) NAAPS (Lynch et al. 2016) and the Météo-France Modèle de Chimie Atmospherique à Grande Echelle (MOCAGE) (Guth et al. 2016). A summary of the current status of global NWP efforts on aerosol forecasting is provided by Xian et al. (2019). Among these efforts, the ECMWF IFS system is considered a strongly coupled aerosol-atmosphere data assimilation system because a single data assimilation algorithm is employed to update both aerosol and atmospheric states (Benedetti et al. 2009). Although the JMA MASINGAR is an inline forecast model of aerosol that is coupled to an atmospheric model, data assimilation of aerosol into MASINGAR is performed separately from the atmospheric data assimilation (Yumimoto et al. 2018).

    In addition to operational efforts, numerous research efforts have addressed the assimilation of aerosol and/or chemistry data into research forecast models for the improvement of weather and air quality simulations (Collins et al. 2001; Weaver et al. 2007; Wang and Niu 2013; Zhang et al. 2014; Lee et al. 2017; Eltahan and Alahmadi 2019). Among them, U.S. National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) provides global reanalysis dataset of both atmospheric and aerosol fields using their GEOS-5 (Randles et al. 2017). Unlike GEOS-5, the WRF-Chem (Grell et al. 2005), which is developed and maintained by the National Center for Atmospheric Research (NCAR), is a widely used research model for regional aerosol, air quality, and atmospheric studies. Similar to WRF-Chem, the RAMS model is also a research model developed for studying regional aerosol-atmosphere interactions.

2.2 Global Versus Regional Applications

Unlike global applications, specifying realistic lateral boundary conditions is critical to regional simulations and data assimilation, in general (Chikhar and Gauthier 2017). A study by Tang et al. (2009) examined the impact of specifying lateral boundary conditions from six different sources on the simulation of tropospheric ozone over the continental U.S., which include a fixed ozone profile, three time-varying ozone profiles derived from global models, and two time-varying ozone profiles derived from soundings. Their results suggest that specifying lateral boundary conditions

with those derived from global models improves the simulation most significantly; however, they found that uncertainties associated with the global models can also translate to the corresponding regional simulations. In addition, Chikhar and Gauthier (2017) pointed out that biases can emerge from the differences in spatial resolution as well as physical parameterizations used between the regional model and the global model, which provides lateral boundary conditions for the regional simulations. Such an issue can be reduced by using a unified system where a regional model and its global version are used together to provide lateral boundary conditions.

3 Aerosol Observation and Forward Operator

3.1 Retrievals Versus Direct Measurements

For analyses and therefore model forecasts to benefit from coupled aerosol-atmosphere assimilation, aerosol observations must be available similarly to atmospheric variables. These observations generally fall into two categories: direct assimilation of aerosol-affected satellite radiances or the assimilation of retrieved aerosol products. Both approaches carry distinct strengths and weaknesses. For example, direct assimilation would necessitate complex radiative transfer code which would lead to costly computational time. On the other hand, retrieved observations inherently make assumptions related to the physical characteristics of aerosols. These include species type, shape, size (bulk or binned categorization), and refractive indices. With that, retrieved products must then be matched to a particular model. Even with these challenges assimilation of retrieved aerosol products is the current operational approach as it affords the availability of quality observations with estimates of uncertainty. The following subsections briefly describe currently available aerosol observations.

3.1.1 Aerosol Optical Depth

An example of a retrieved aerosol product is the aforementioned AOD. As the name suggests, AOD is a quantity that measures the loss of light due to scattering and absorption through a vertical column. This quantity depends on the type and physical characteristics of the aerosols that are present. Ground-, airborne-, and spaceborne-based AOD observations have been used in a variety of data assimilation systems (variational, ensemble, hybrid) at National Center for Environmental Prediction (NCEP), ECMWF and NRL. Liu et al. (2011) showed that 3D-Var assimilation of AOD from the Moderate Resolution Imaging Spectroradiometer (MODIS) improved both aerosol analyses and aerosol forecasting. Further, Benedetti et al. (2019) utilized 4D-Var to assimilate MODIS AOD observations and demonstrated improvement in dust analyses and forecasts for up to 48 h in East Asia. Examples

of ensemble-based assimilation of aerosols can also be seen in Pagowski and Grell (2012), Rubin et al. (2016), and Schwartz et al. (2014). Hybrid data assimilation has also been shown to be effective in aerosol analyses and forecasts (Schwartz et al. 2014; Choi et al. 2020).

3.1.2 Satellite Radiances Affected by Aerosols

Visible, ultraviolet (UV) and near-infrared wavelengths could very well be the future of aerosol assimilation. This has been shown possible (Weaver et al. 2007) but several challenges have prevented this from becoming operationally viable. These include the speed and complexity of the available radiative transfer codes, complexity of the model, and how polarization would be addressed. A benefit of direct assimilation would be the ability to assimilate from different satellite instruments. Currently attempts are underway at ECMWF to assimilate two aerosol visible radiances from MODIS and have been shown to be effective in representing plumes in the 4D-Var analyses comparable to the available observations. While direct assimilation of aerosol-affected satellite radiances has been shown viable, future research is still required for this to become operationally feasible.

3.1.3 LIDAR

Light Detection and Ranging (LIDAR) instruments use a pulsed laser to generate three-dimensional observational imagery of the Earth's atmosphere and surface characteristics. This is done by observing the backscatter from molecules and particles. LIDAR instruments can retrieve profiles describing the composition of atmosphere in regard to water content and aerosols and also determine wind fields. One such example is the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite. CALIPSO utilizes LIDAR along with infrared and visible imagers to observe clouds and aerosols and is part of the "A-Train" satellite constellation. The vertical profiles retrieved by CALIPSO have been able to provide highly-accurate cloud heights and high thin cirrus clouds which have been difficult to observe previously.

Given the utility and quality of LIDAR observations, there are other space-borne instruments that are either in the pre-launch design phase or that have recently been launched and are now used operationally. The EarthCARE satellite is part of the European Space Agency's Earth Explorer Programme and scheduled to launch in 2022. EarthCARE will carry LIDAR, radar, radiometers, and imagers with the goal of producing high-resolution horizontal and vertical profiles of aerosols, liquid water, cloud distribution, and atmospheric radiative heating and cooling. These new datasets of highly-variable parameters are expected to improve forecasting and climate modelling.

Aeolus, another space-borne satellite, was launched in 2018 and has been used operationally at ECMWF since January 2020. Aeolus employs a LIDAR instrument

capable of observing the Doppler shift of atmospheric molecules and particles to retrieve highly precise wind profiles. While the wind profiles are currently used to improve numerical weather prediction, Aeolus also has the ability to retrieve aerosol optical properties such as extinction and optical depth. The value and use of these aerosol profiles has yet to be fully explored.

3.1.4 AERONET

The AERONET (AErosol RObotic NETwork) program is a network of ground-based sun photometers capable of measuring atmospheric aerosol properties. By measuring sun and sky radiances at a fixed number of wavelengths in the visible and near-infrared spectrum, precipitable water and aerosol properties such as AOD, single scattering albedo, aerosol scattering phase function, and aerosol volume size distribution can be retrieved. This global network has grown to over 600 sites as of 2018. AERONET thus provides a vast database of ground-truth calibration data for current and future satellite instruments which is a crucial component of utilizing new observations in data assimilation to improve numerical weather prediction.

3.2 AOD Observation Operator

To assimilate AOD observations a data assimilation system must include a forward operator (also known as observation operator) that computes a model-equivalent value of AOD. This operator will be unique to the numerical prediction model as it depends on the represented aerosol species. Each aerosol species has specific physical properties including effective radius and wavelength-dependent indices of refraction. These characteristics must be known to calculate the mass extinction coefficient via Mie theory (Bohren and Huffman 1983). To account for hygroscopic growth -Köhler theory, Petters and Kreidenweis (2007) grows each particle to equilibrium per the ambient relative humidity. Since the Mie calculations can be expensive, look-up-tables can be created offline for quick reference of a species' humidity-dependent mass extinction coefficient (Eq. 4). This technique has been applied in this study (see Sect. 5.3). Total-column AOD is then computed by summing over all species and model levels following Liu et al. (2011) and Pagowski et al. (2014). The calculation of AOD at a given wavelength λ (nm) is expressed as

$$AOD(\lambda) = const \cdot \sum_{i=1}^{N_{aero}} \sum_{k=1}^{k_{top}} E_{ext}(\lambda, n_{ri}, r_{effi}) \cdot c_{ik} \cdot \frac{\Delta p_k}{g} \qquad (4)$$

where $AOD(\lambda)$ represents the spectrally dependent AOD operator (unit less), i is the index for aerosol species, N_{aero} is the total number of aerosol categories that contribute to the AOD calculation, k is the index for model vertical levels, and k_{top} is

the model top level. E_{ext} is the spectrally dependent mass extinction coefficient (m^2 g^{-1}), which is a function of the index of refraction n_r and effective radius r_{eff} (nm) of a given aerosol species, c_i, in the form of mass mixing ratio (g of aerosol/kg of dry air). Δp_k is the pressure difference (mb) between two vertical levels k and $k + 1$, and g is the acceleration due to gravity (m s^{-2}). *const* is a constant of 10^5, as a result of unit conversion (Eq. 4).

3.3 AOD Error and Bias Estimation

AOD observations include both a quality flag and a definition of an observational error which depends on the retrieval algorithm, e.g. MODIS Dark Target (Levy et al. 2013) versus Deep Blue (Hsu et al. 2006). Ideally these definitions would extend to error covariances which would describe correlations between different aerosol products in both space and time. Moreover, to improve assimilation of these observations, an estimation of bias and the ability to correct for it, is also desired. These bias correction procedures can be generally categorized as either static (offline) or variational. The static bias correction scheme (Eyre 1992) considers differences in the observations and the model state over a period of time and defines bias predictors using satellite scan angle along with several atmospheric variables (e.g. skin temperature, total column water, etc.). This is carried out offline for each satellite sensor and band and is frequently updated. The bias correction is then applied to the observations in the data assimilation system. Variational bias correction methods include bias coefficients within the state vector of the minimized cost function. Therefore, these coefficients are continuously updated, along with the state vector itself, during each data assimilation cycle. The bias is defined as a linear combination of predictors, similar to the static scheme, using scan angle along with atmospheric variables. More details can be found in (Derber et al. 1991; Parrish and Derber 1992; Derber and Wu 1998; Dee 2005; Auligné et al. 2007).

4 Challenges

4.1 Choice of Control Variables

The choice of control variables is directly related to the background error covariance, which plays a fundamental role in data assimilation. Control variables can be defined as a subset of variables of an NWP system that can potentially impact its prediction. In general, control variables include not only the initial conditions of prognostic variables of an NWP system, but also non-prognostic variables, empirical model parameters, and model error bias. The particular choice of control variables depends on both the feature of interest (e.g., tropical cyclones, thunderstorms, and blowing

dust) and the type of observed data to be assimilated (e.g., satellite radiances, radar data, and satellite retrieved quantities). Over the previous decades, research efforts focused on improving the forecast of severe thunderstorms, as a result, doppler radar data was assimilated; therefore, a choice of control variables would be the horizontal component of the prognostic wind (e.g., Sun 2006; Hu and Xue 2007). Another example of a feature of interest is blowing dust. Progress has also been made in the assimilation of airborne dust using aerosol-atmosphere coupling.

An important aspect of a strongly coupled aerosol-atmosphere data assimilation system is to have a set of control variables that cover the state of both components (Pagowski et al. 2014). Control variables associated with the aerosol component include the initial conditions of dust, sea salt, carbon particles from agricultural or wildfire burning, and sulphate from agricultural and industrial sources, which are some of the typical aerosol species. In addition, control variables associated with the atmospheric component typically include the initial conditions of temperature, pressure, all components of the three-dimensional wind vectors, and water vapor mixing ratio. Having a set of control variables, which covers both components, allows the information of assimilated observations to be spread into relevant variables via the background error covariance matrix, which includes cross-component correlations (to be discussed in Sect. 4.2).

Additional information is now given to a specific scenario: airborne dust, which results from high winds over semi-arid surfaces. Similar to cloud microphysical schemes, aerosol solvers/models predict moments of aerosol species. In particular, a single-moment scheme predicts only the mass mixing ratio (ug of dust per kg of air) of a given aerosol species (e.g., WRF-Chem; Grell et al. 2005); whereas a double-moment scheme predicts both mass mixing ratio and number concentration (number of dust particle per kg of air) of a given aerosol species (e.g., RAMS aerosol module; Saleeby and van den Heever 2013). Consequently, a double-moment scheme allows three-dimensional variability in particle size, because particle size is a function of both mass and number concentration. Efforts to advance the field of dust assimilation have focused on the first moment, the mass field, as a first step. After the assimilation of dust with the first moment becomes better understood, the next step is to include the second moment, number concentration, in the set of control variables.

There are important challenges when including only the first moment as a control variable. As a result of altering only the first moment, during the data assimilation process, mass may appear in a region devoid of number concentration. As previously stated, particle size depends on both moments. Consequently, if there is a region in a numerical domain with non-zero mass and non-existence number concentration, then calculation of particle size becomes problematic. An additional challenge is that a forecast from an analysis, which contains an inconsistency between the two moments, will cause numerical errors. Although focus was placed on the first moment (mass) and the second moment (number concentration), the above discussion applies equally well to any double-moment prognostic variables like the first and second moments of cloud microphysics (Cotton et al. 2003; Saleeby and Cotton 2004).

In preparation for a discussion of background error covariance (Sect. 4.2), additional care should be exercised in choosing control variables. Because the background

error covariance matrix is computed from the set of control variables, choice of the set of control variables has fundamental impact on the efficiency and success of assimilation (Xie and MacDonald 2012; Sun et al. 2015).

4.2   Background Error Covariance

As mentioned in the previous section, background error covariance matrix provides a mechanism for spreading the information from assimilated observations to control variables represented by grid points (both horizontally and vertically) of all coupled components (Fisher 2003). In addition, background error covariance not only allows observation of different types to act in synergy, but also helps maintain the analysis state closer to balance (Bannister 2008a). Having chosen a set of control variables does not naturally guarantee a corresponding background error covariance matrix that can accurately represent the associated actual error. Careful tuning and possibly modeling of background error covariance is required for any effective data assimilation schemes that include variational, ensemble, as well as hybrid methods.

Due to its prohibitive size (NWP system has a large dimensional state space ~ 10^8), the use of the explicit form of background error covariance matrix is impossible. Instead, several techniques have been developed to measure characteristics of background error statistics for modeling and specifying realistic background error covariance matrix. A review of measuring and modeling background error covariance in the context of atmospheric data assimilation systems was provided in Fisher (2003) and Bannister (2008a, 2008b). Methods to measure the background error statistics include the following: analysis of innovations, differences between forecasts of different lengths that verify at the same time (i.e., the National Meteorological Center (NMC) method; Parrish and Derber 1992), the lagged NMC methods, and the ensemble-based Monte Carlo method. In particular, the NMC method is widely used by several NWP centers due to its advantage of low computational cost. However, the NMC method was often found to overestimate covariances due to the use of longer forecast lengths, e.g., 24 h and 48 h, to estimate errors of the background, which is usually a 6 h forecast. Following the measurement of background error statistics, the modeling of background error covariance can be achieved via spectral/wavelet methods (Fisher 2006) and control variable transform (Bannister 2008b), both of which seek to simplify the representation of the background error covariance matrix and were developed for variational-based schemes.

Benedetti and Fisher (2007) and Kahnert (2008) were the first to apply the NMC method to estimating background error statistics of aerosols. With that, a satisfactory background error covariance matrix was constructed with the use of a wavelet modeling approach without the need to prescribe the vertical and horizontal correlation (Benedetti et al. 2009). In addition, a generalized background error covariance matrix model was developed by Descombes et al. (2015) as a community tool to be used beyond atmospheric applications (e.g. geophysical, chemistry, etc.).

In the context of ensemble data assimilation, background error covariance can be created with the use of ensemble of model forecasts. As such, the ensemble background error covariance matrix is time-dependent and includes embedded correlations between control variables from the model. Nevertheless, additional care is still required to fine tune the ensemble background error covariance to avoid filter divergence as well as spurious correlation due to the use of a much smaller ensemble (i.e., reduced rank). In general, a good practice to visualize the structure functions of the background error covariance (Thépaut et al. 1996) can be achieved via examining analysis increments of control variables resulted from assimilating a single observation of the kind of a control variable in a pre-specified grid point (i.e., single observation experiment; see Sect. 1.3).

4.3 Non-Gaussianity and Non-Linearity

Many variational and ensemble-based data assimilation and retrieval systems assume that the observational and model errors come from a Gaussian distribution. Previous research has indicated this is not necessarily true for variables that are not from a Gaussian distribution, e.g. variables that are positive definite such as humidity or total precipitable water. Recent research has sought to address this limitation by introducing a cost function based on a mixed Gaussian-lognormal distribution (Fletcher and Jones 2014). Here the incremental 3D and 4DVAR formulations of the mixed distribution cost function is derived and improved performance is shown with experiments based on the Lorenz 1963 toy model. This formulation has also been shown to improve 1DVAR water vapor mixing ratio retrievals (Kliewer et al. 2016) as this variable is certainly positive-definite. Another recent approach that avoids any assumption of probability distribution is with the application of particle filters (Van Leeuwen 2010) however these methods have not been found to be operationally viable as of yet due to their computational cost.

The non-Gaussian nature of AOD can certainly have an impact on the quality of the coupled data assimilation. As previously described the forward operator for AOD observations is certainly non-linear since it incorporates hygroscopic growth as a function of relative humidity. Preliminary experiments have confirmed this by noting that the distribution of innovations during assimilation is often positively or negatively skewed. While this issue can have impact on the data assimilation analyses and the subsequent NWP forecasts, this is out of the scope for what is presented here and is not addressed within these experiments.

4.4   Insufficient Data for Independent Verification

A standard way of measuring the success of data assimilation performance is to compare its analysis and background in observation space against independent observations, i.e., observations not assimilated. Benedetti et al. (2018) describe several observation types that can be used for verifying chemistry and aerosol data assimilation. However, there are situations in which the number of observations available for assimilation is limited and/or their representativeness is inadequate (a few pointwise observations to validate global—over all points—data assimilation). This is especially relevant to aerosol data assimilation, and in particular to regional aerosol data assimilation. Commonly used verification data include AERONET and CALIPSO. Although proven useful, there are some concerns when using them for verification of aerosol data assimilation, and in particular regional aerosol data assimilation, related to their limited spatial and temporal coverage. Given that typical Gaussian data assimilation involves some kind of optimization over all grid points and observations, having a few pointwise observations such as AERONET is not sufficient for verifying data assimilation. Similar is true for CALIPSO, which produces a high-resolution but narrow-swath vertical cross-section of aerosol.

One can also think of additional issues that may become important. For example, a new satellite sensitive to a particular aerosol variable that is rarely observed is launched with a goal of demonstrating the usefulness of new observation type in data assimilation. Under these assumptions there are likely no other, independent observations similar to the new satellite and therefore direct verification is not possible. Another example may be the limitation introduced by choosing the area of interest that is sporadically observed, such as polar regions, oceans, and deserts. Without sufficient statistically independent observations such studies may never be properly verified. Although a particular research may be of great scientific interest, not having independent observations to verify data assimilation performance could preclude efforts to assimilate these observations.

Described scenarios may be more common in regional data assimilation applications, but they could happen in global applications as well. This is because observation operators that transform control variables to observed variables often only have local impact, especially in the horizontal directions. Imagine a case when a special type of observation is available and assimilated only over a small area of a global domain while the verifying independent observations are not available, it will not be possible to reliably assess the impact of assimilated data.

All of the above suggests that there is a need to address alternative verifications for data assimilation in general, and particularly for aerosol, without using independent observations. The main underlying premise of such an approach is that a data assimilation algorithm contains additional information that is overlooked and consequently not used for its verification.
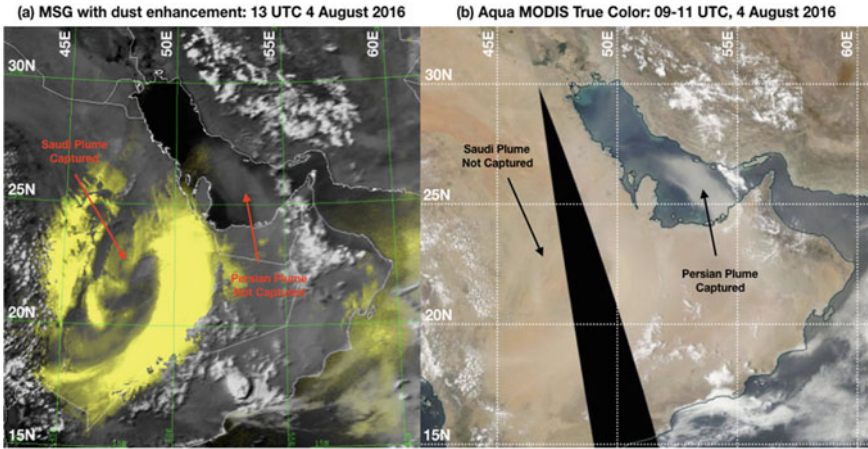
Fig. 3 Satellite imagery of the two dust plumes over the Arabian Peninsula on 4 August 2016: **a** Meteosat Second Generation (MSG) imagery with dust enhancement applied (showing dust in yellow) and **b** Aqua MODIS true color imagery

5 Experiments and Results

5.1 Case Study

A dust storm case over the Arabian Peninsula, one of the major dust sources of the world and the so-called dust belt (Jish Prakash et al. 2015), occurred on 4 August 2016 (Miller et al. 2019; Saleeby et al. 2019) and was chosen to illustrate the utility of a strongly coupled aerosol-atmosphere data assimilation system. On 4 August 2016, two distinct dust plumes occurred (Fig. 3), in which one plume advected offshore of the United Arab Emirates (UAE) to the central portion of the Persian Gulf (referred to as the Persian Plume; Fig. 3a), which was detected by the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard Meteosat Second Generation (MSG) - 8 imagery with dust enhancement algorithm applied, and the other plume was located in interior regions of Saudi Arabia (referred to as the Saudi Plume; Fig. 3b), which was detected by Aqua MODIS true color imagery. As discussed in Miller et al. (2019), the environment of the Saudi Plume was characterized by values of total precipitable water (TPW) less than approximately 25 mm whereas the Persian Plume was in an environment characterized by values of TPW in excess of 45 mm.

5.2 Overview of the RAMS-MLEF System

In order to demonstrate the utility of a strongly coupled aerosol-atmosphere data assimilation system, an NWP model was interfaced to a data assimilation system.

That is, RAMS (Cotton et al. 2003) was interfaced with the Maximum Likelihood Ensemble Filter (MLEF; Zupanski 2005; Zupanski et al. 2008), hereafter referred to as the RAMS-MLEF system, to conduct experiments for the 4 August 2016 case. Before the experimental setup is described, a brief introduction to RAMS, MLEF, and the RAMS-MLEF system is provided.

RAMS is a multi-purpose mesoscale numerical prediction model that was developed at CSU. Throughout the years, RAMS has undergone multiple upgrades that include improvements to its microphysics via the implementation of a bimodal and double-moment cloud water scheme (Saleeby and Cotton 2004), an improved capability to assimilate lightning data (Federico et al. 2017), and the development of an interactive aerosol module (Saleeby and van den Heever 2013). Of these recent upgrades, the development of a RAMS aerosol module is directly related to the study herein. There are a total of nine aerosol categories represented by the aerosol module in RAMS: (i) submicrometer sulphate, (ii) supermicrometer sulphate, (iii) submicrometer mineral dust, (iv) supermicrometer mineral dust, (v) film-mode sea salt, (vi) jet drop-mode sea salt, (vii) spume-mode sea salt, (viii) submicrometer regenerated aerosols, and (ix) supermicrometer regenerated aerosols. For each aerosol category, the size is represented by a lognormal distribution given by

$$n(r) = \frac{N}{r\sqrt{2\pi}\,ln(\sigma_g)}exp\left[-\frac{\left(ln\frac{r}{r_g}\right)^2}{2ln^2\sigma_g}\right] \tag{5}$$

where $n(r)$ is number concentration of aerosols of dry radius r, N is total number concentration of aerosols, r_g is lognormal distribution geometric median radius, and σ_g is lognormal distribution geometric standard deviation. Although the shape of the size distribution as described in Eq. (5) is fixed during a simulation, the distribution is allowed to translate in the direction of r. That is, as a result of sources and sinks of aerosol mass during a simulation, the size distribution given in Eq. (5) is allowed to shift toward larger or smaller values of r. In addition, the width of the size distribution is determined by σ_g, which behaves like a dispersion parameter in a Gamma size distribution used in microphysical development.

MLEF is a hybrid data assimilation algorithm with both variational and ensemble features. Similar to other data assimilation methods (e.g., Evensen 1994; Houtekamer and Mitchell 2001; Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002), a generalized flow chart of MLEF also consists of a forecast step and an analysis step. During the forecast step, MLEF generates an ensemble of forecasts to estimate the flow-dependent background/forecast error covariance. After completion of the forecast step, minimization of a prescribed cost function occurs during the analysis step, see Fig. 4, where **x** and **y** represent the state vector and the observation vector, respectively; subscript **f** denotes the forecast (or background) and subscript **a** denotes the analysis; $\mathbf{P_f}$ is the flow-dependent background/forecast error covariance matrix and $\mathbf{P_a}$ is the analysis error covariance matrix; superscript **t** denotes time; h denotes a collection of observation operators; m represents a forecast model. Unlike pure
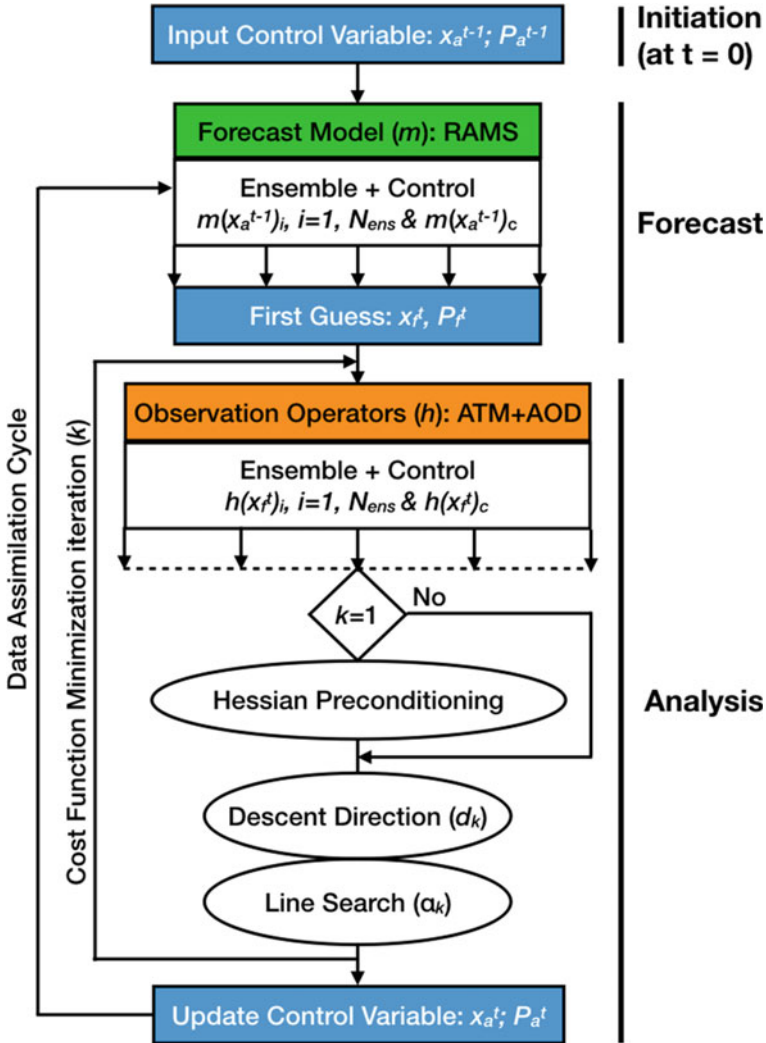
Fig. 4 A flow chart of the RAMS-MLEF system. Interfaces between MLEF and RAMS are highlighted in the following colors: blue boxes represent interfaces for Input/Output (I/O) between MLEF and RAMS, the green box represents the interface as a driver to call and run RAMS, and the orange box represents the interface for observation operators, which require input from RAMS

variational methods (e.g., Parrish and Derber 1992; Zupanski 1993; Rabier et al. 1999), MLEF, a hybrid system, solves the prescribed cost function, Eq. (6), with Hessian preconditioning in the ensemble space,

$$J(x) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x_b})^T P_b^{-1}(\boldsymbol{x} - \boldsymbol{x_b}) + \frac{1}{2}[y - h(x)]^T R^{-1}[y - h(x)] \qquad (6)$$

where **R** is the observation error covariance, which is often a diagonal matrix following the assumption that observations are not spatially correlated. Although any forecast model, as indicated by *m* in Fig. 4, can be interfaced with MLEF, this study utilizes RAMS.

A schematic diagram shown in Fig. 4 outlines the components of the RAMS-MLEF system. Specifically, three interfaces are implemented in MLEF and they are (1) Input/Output (I/O) interfaces between MLEF and RAMS, (2) an interface that acts as a driver to call and run RAMS, and (3) an interface for observation operators that utilize RAMS output to compute the first guess of assimilated quantities as part of the innovation of data assimilation. In MLEF, observation operators for atmospheric observations are adapted from the forward component of the Gridpoint Statistical Interpolation (GSI; Wu et al. 2002; Kleist et al. 2009) through a module as illustrated by ATM in the orange box of Fig. 4. With that, atmospheric observations that are provided by NCEP, such as the conventional observations (e.g., radiosonde, surface station, buoy, etc) within the NCEP Prepared Binary Universal Form for the Representation of meteorological data (PrepBUFR) dataset and non-conventional atmospheric observations provided by satellite radiances data from various platforms, can be assimilated by MLEF, which is consistent with operations at NCEP. However, the AOD observation operator that is embedded in the Community Radiative Transfer Model (CRTM; Han et al. 2006), which is one of the observation operators within GSI, was specifically designed for the GOCART (Chin et al. 2000) aerosol species. Similarly, an AOD observation operator was developed specifically for the RAMS aerosol module within the RAMS-MLEF system.

In the RAMS-MLEF system, an observation operator for AOD specific for the RAMS aerosol module was developed in accordance with Eq. (4). Out of the nine aerosol categories, eight of them are used, i.e., $N_{aero} = 8$, to calculate AOD for this study. Supermicrometer sulphate is not used due to its little contribution to the total AOD. The optical properties of the eight aerosol categories at 0.55 μm under dry conditions are provided in Table 1. The mass extinction coefficient is computed using Mie theory, in which the spherical assumption of aerosol particles is required. For each of the aerosol categories, particles are first grown hygroscopically to equilibrium with ambient relative humidity using κ-Köhler theory (Petters and Kreidenweis 2007) and the refractive index is adjusted based on volume mixing with water. To reduce computational expense, a lookup table of the mass extinction coefficient as a function of ambient relative humidity (RH, %) for each of the eight aerosol categories at 0.55 μm is prepared. A 1% interval of RH is used in the lookup table, which is plotted in Fig. 5. For a simulated RH with a value that falls between two integer numbers (e.g., 85.6%), the integer value that is closer to the simulated value will be used (e.g., 86%).

Configuration of the RAMS-MLEF used for this study is now described. A time-lagged methodology (Suzuki and Zupanski 2018) is used to generate an initial set of N ensemble RAMS forecasts, which are valid at a prescribed initial time (0000 UTC 03 August 2016 is used for this study). As mentioned in Suzuki and Zupanski (2018), the so-called time-lagged methodology involves running a single deterministic or control forecast centered at the initial time ($t = 0$) of data assimilation, i.e., from $t =$

Table 1 Optical properties for the RAMS aerosol categories under dry conditions and their hygroscopicity parameters

| Aerosol categories	Geometric median diameter (μm)	Effective radius (μm)	Geometric standard eeviation (μm)	Density (g cm^{-3})	Index of refraction at 0.55 μm (real)	Index of refraction at 0.55 μm (imaginary)	Mass extinction coefficient (m^2 g^{-1}) at 0.55 μm	Hygroscopicity parameter (kappa)
sub-micrometer sulfate	0.08	0.095	1.8	1.76	1.524	1.00E−07	2.44	0.55
sub-micrometer mineral dust	1.398	1.658	1.8	2.5	1.5	0.0005	0.411	0.04
super-micrometer mineral dust	5.9	6.997	1.8	2.65	1.5	0.0005	0.0854	0.05
film-mode seasalt	0.2	0.237	1.8	2.2	1.5	1.00E−08	3.04	0.8
jet drop-mode sea salt	2.0	2.372	1.8	2.2	1.5	1.00E−08	0.312	0.8
spume-mode sea salt	12.0	14.232	1.8	2.2	1.5	1.00E−08	0.0514	0.8
sub-micrometer regenerated aerosol	0.02	0.024	1.8	2.4	1.524	6.00E−03	0.166	0.17
super-micrometer regenerated aerosol	2.0	2.372	1.8	2.4	1.524	6.00E−03	0.298	0.17
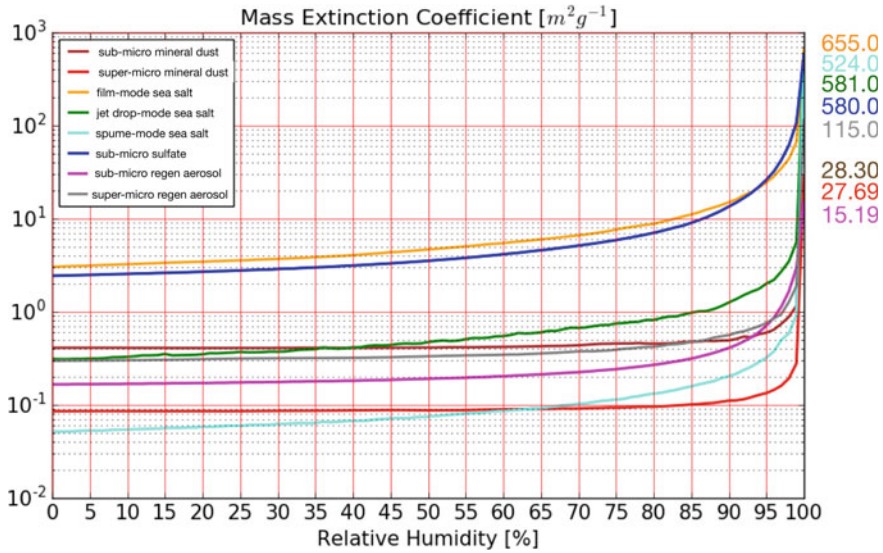
Fig. 5 Mass extinction coefficient (m^2 g^{-1}) as a function of relative humidity (RH ; %) at 0.55 μm for the eight RAMS aerosol categories listed in Table 1. Colored numbers on the right-hand side of the figure indicate values of mass extinction coefficient at RH $= 100\%$

$-T$ to $t = +T$, where T is a specified assimilation window ($T = 6$ h is used in this study). During this deterministic forecast, RAMS is configured to generate output at every *2 T/N* step and thus creating $N + 1$ output, where N denotes the size of the ensemble ($N = 32$ for the August 2016 study). Out of the total of $N + 1$ output, the output that is valid at $t = 0$ is denoted by an Mx1 column matrix x_c, where M is the total number of control variables times grid points of a RAMS domain and c indicates the control member. The other N outputs are used to define ensemble perturbations (p_i, $i = 1$, N) at $t = 0$ by calculating $p_i = \frac{1}{\sqrt{N}}(x_i\text{-}x_c)$, where x_i is the state from an ensemble member and p_i is one column of a matrix whose square is $\mathbf{P_f}$.

Each assimilation cycle of the RAMS-MLEF system begins with a 6 h ensemble and control forecasts and ends with a control analysis along with the associated analysis error covariance, $\mathbf{P_a}$. At the end of the ensemble and control forecasts of any cycle, $\mathbf{P_f}$, which contains the cross-component ingredients for strongly coupled data assimilation, is re-computed and used as part of the cost function for the assimilation of observational data. Results at the end of a cycle include an updated x_c, i.e., the analysis field, and the associated analysis error covariance, which is used to characterize the uncertainty of the analysis field.

Covariance inflation is used to increase the ensemble spread during each assimilation cycle. Due to the use of identical lateral boundary conditions, ensemble members may collapse. One way to avoid ensemble members from collapsing is to use the covariance inflation methodologies described in Zhang et al. (2004) and Whitaker

and Hamill (2012), which act to increase the ensemble spread in order to account for unrepresented error of sources. In the RAMS-MLEF system, a linear combination of these two methods are used, where 50% of weight is given to the method described in Whitaker and Hamill (2012) and 50% of weight is given to the method described by Zhang et al. (2004).

As mentioned earlier, success of a coupled data assimilation system is highly dependent on the choice of control variables. A set of control variables used in the RAMS-MLEF system includes the following: the three-dimensional wind components (u, v, and w), perturbation Exner function (pi), ice-liquid water potential temperature (θ_{il}), water vapor mixing ratio (rv), and the mass mixing ratio of the sub- and super-micrometer mineral dust (*md1mp* and *md2mp*). Because RAMS uses a leapfrog time stepping scheme, two temporal solutions, t_1 and t_2, exist only for the u, v, w, and *pi* prognostic variables, where an Asselin filter (Cotton et al. 2003) is used to prevent the two temporal solutions from diverging via damping the computational mode. In order to preserve the difference of the two temporal solutions for u, v, w, and *pi*, the RAMS-MLEF system stores the differences before assimilation occurs, and then only alters the t_1 solution of u, v, w, and *pi* during the assimilation. After assimilation, the t_2 solution will be updated through the use of the stored differences. As a consequence, the differences between t_1 and t_2 stay the same before and after data assimilation even through both time solutions are changed. Note that the RAMS aerosol module (Saleeby and van den Heever 2013) uses a double-moment scheme, which predicts both mass mixing ratio and number concentration for all 9 aerosol categories.

As stated in Sect. 4.1, prediction of both mass (first moment) and number concentration (second moment) of dust may be included into a data assimilation study. Mass and number concentration for both the sub- (*md1mp and md1np)* and super-micrometer mineral dust (*md2mp and md2np*) are predicted by the RAMS aerosol module. Dust mass and numbers are predicted for two different particle sizes; one for the sub-micrometer ($\sim 0.41\,\mu$m radius) mineral dust, second for the super-micrometer ($\sim 1.74\,\mu$m radius) mineral dust. In other words, mass and number for sub-micrometer (super-micrometer) mineral dust is referred to as dust bin 1 (dust bin (2)). As stated above, only mass in each dust bin is updated during assimilation of observed quantities of dust, which results in an inconsistency between dust mass and numbers for each dust bin of an analysis. One method to rectify the inconsistency between mass and number in an analysis is to assume an average dust particle size for each dust bin and recompute the number concentration of each dust bin from the updated mass field and assumed particle size. Consequently, both mass and numbers in each dust bin within an analysis become consistent with one another. Since u, v, w, *pi*, θ_{il}, *rv,* and both moments of each dust bin have been updated, the next assimilation cycle begins with the forecast initialized from the analysis.

5.3 Application of the RAMS-MLEF System

One RAMS-MLEF experiment named ATMAOD is carried out from 0000 UTC 03 August to 0600 UTC 04 August 2016 with a 6-hourly data assimilation cycle (total of 6 cycles). In this ATMAOD experiment, both the conventional atmospheric observations from NCEP PrepBUFR dataset and the 0.55 μm MODIS AOD retrievals are assimilated. There is only one domain used and the domain is composed of 400 east–west, 225 north–south, and 50 vertical grid points. In Fig. 6, the NCEP PrepBUFR dataset used in the ATMAOD experiment is displayed. Note that the majority of the dataset is only available at the surface and is indicated by green, blue, and orange symbols. Red symbols indicate the location of rawinsondes, which are the only source of conventional data that provide information from the surface to approximately the lower stratosphere of the atmosphere.

Due to the availability of MODIS data that is used to produce AOD retrievals, AOD retrievals are only assimilated at the cycle 2 (0600 UTC 03 August), cycle 3 (1200 UTC 03 August) and cycle 6 (0600 UTC 04 August) of the ATMAOD experiment (Fig. 7a). For the study herein, an observation error value for the AOD retrievals, a unitless quantity, is 0.1. Similar to Remer et al. (2005) and Liu et al. (2011), AOD observation error (*Err*) is increased by 5% (15%) for ocean (land) scenes (see Eq. 7).

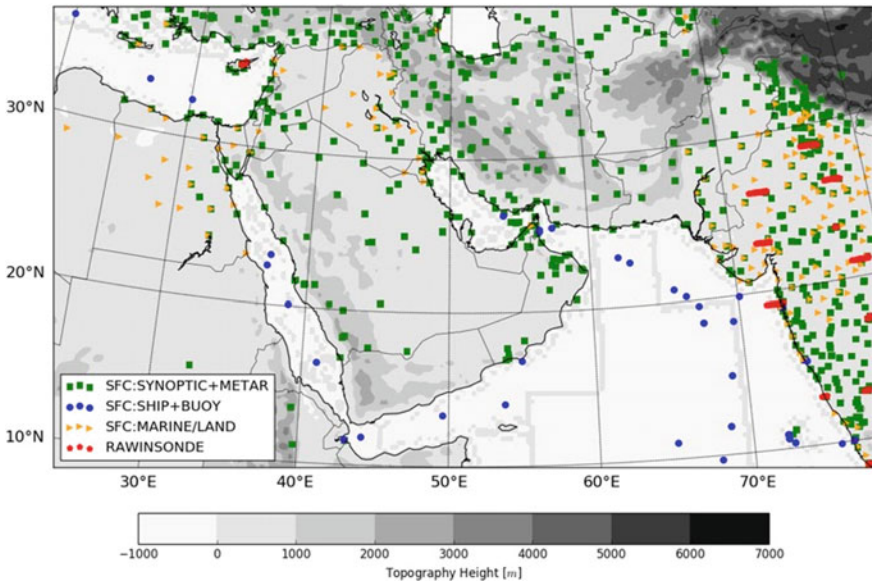$$\text{Err}_{\text{ocean}} = 0.1 + 0.05 * \text{AOD}$$

Fig. 6 NCEP PrepBUFR dataset that was assimilated into the ATMAOD experiment over the RAMS domain that covers the Arabian Peninsula. Topographic height (m) is plotted in gray scale

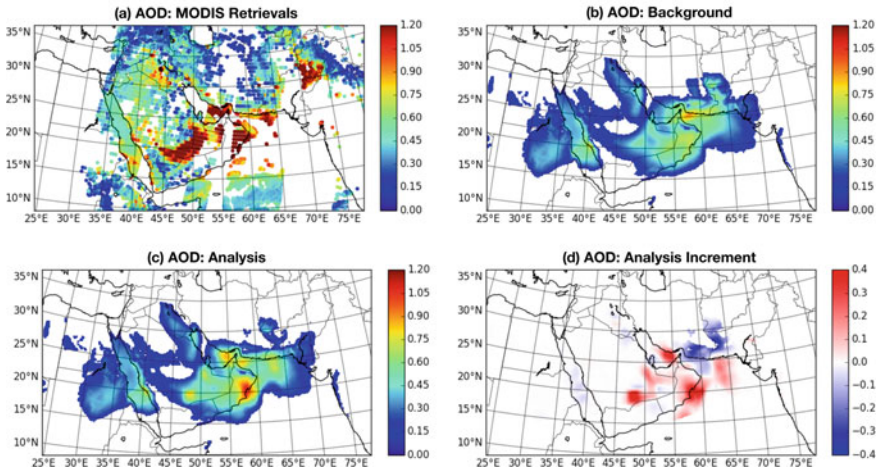### RAMS-MLEF ATMAOD Experiment valid at 06 UTC 4 August 2016

Fig. 7 Horizontal distribution of AOD: **a** retrievals from MODIS and **b–d** RAMS simulated AOD field computed from cycle 6 of the RAMS-MLEF ATMAOD experiment: **b** background, **c** analysis, and **d** analysis increment, i.e., analysis minus background: **c–b**. Note that AOD is a unitless quantity. Valid time is 0600 UTC 4 August 2016

$$\text{Err}_{\text{land}} = 0.1 + 0.15 * \text{AOD} \tag{7}$$

In order to reduce the effects of spatially correlated observation error, data thinning is applied to the AOD retrievals prior to the actual assimilation. For a given cycle, AOD retrievals are first thinned such that every fifth pixel of a given retrieval image is excluded from assimilation and used for verification. Once spatial thinning is completed, the next step is quality control. During the quality control procedure, the so-called gross check is applied to remove large differences (usually three times the prescribed observation error, where observation error is one standard deviation for the assumed Gaussian distribution) between the AOD retrievals and the first guess.

In Fig. 7, assimilated MODIS AOD retrievals (thinned and passed quality control) are presented along with simulated AOD computed from the background and analysis field of cycle 06 of the ATMAOD experiment along with the difference between the analysis and background AOD (i.e. analysis increment). Both the background and analysis appears to have captured the general distribution of AOD (Fig. 7b, c), however, with slightly smaller magnitude compared to the retrievals (Fig. 7a). Nevertheless, after assimilating the AOD retrievals, the representation of the Persian plume (around 55°E and 26°N) and the Saudi Plume (from 45°E and 18°N to 52°E and 23°N) (see Fig. 3) is improved from the background in the analysis of ATMAOD experiment. The analysis increment of AOD further confirms that by assimilating MODIS AOD retrievals, the magnitude of AOD of both plumes are increased from background to analysis to reflect the assimilation.

In addition to the ATMAOD experiment, an AODONLY experiment, in which only AOD retrievals from MODIS were assimilated, was utilized to examine the role of assimilating atmospheric observations in the RAMS-MLEF system. The AODONLY experiment was performed by running a 6 h forecast from the analysis of cycle 5 of the ATMAOD experiment, and then assimilating AOD retrievals into the 6 h forecast valid at 0600 UTC 04 August 2016 for a resulting AODONLY analysis valid at the same time. Another 6 h forecast was run from the AODONLY analysis and was valid at 1200 UTC 04 August 2016. Differences between the two experiments were examined in order to understand impact of assimilating atmospheric observations on simulated dust; that is, variables from the AODONLY experiment were subtracted from the same variables from the ATMAOD experiment. Specifically, total dust ($md1mp + md2mp$) difference at the lowest model level between the two experiments is shown in Fig. 8. Since there were few atmospheric observations over the region of interest (e.g., the Saudi and the Persian plumes), their impact is limited to Persian Gulf coastal areas. In Fig. 8a, where total dust difference at the cycle 06 analysis is shown, one can notice a positive difference in the southeast part of the Persian Gulf, i.e. an increase of total dust due to assimilated atmospheric observations, and
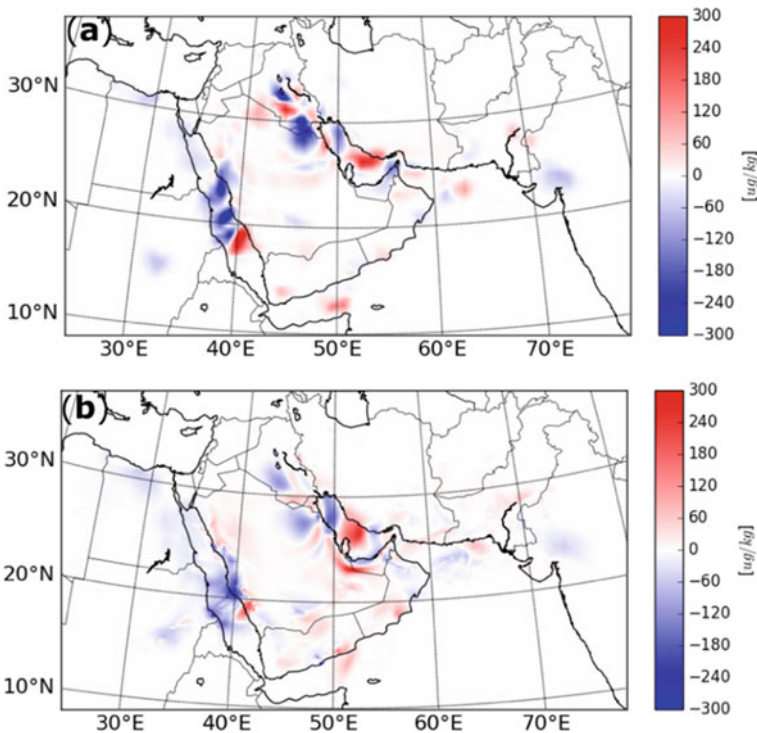

Fig. 8 Total dust (ug kg^{-1}) difference, ATMAOD experiment minus AODONLY experiment, at the lowest model level for **a** the analysis at cycle 6, valid 0600 UTC 4 August 2016 and **b** the 6-h forecast initialized from the analysis valid at 1200 UTC 4 August 2016

a negative difference in the northwest part of the Persian Gulf, indicating a decrease of total dust due to atmospheric observations. A 6 h forecast difference valid at 1200 UTC 04 August 2016 (Fig. 8b) also shows that the analysis differences are generally retained in the forecast. There is subtle change in the magnitude and the pattern of the total dust difference, but it is possible to identify and follow the movement of these changes over the 6 h time period. Such a result indicates that data assimilation was able to transform the information from atmospheric observations to dust initial conditions in such a way that it is supported by coupled model dynamics. More importantly, this result suggests that ensemble cross-covariance in strongly coupled data assimilation can have a satisfactory structure, which is encouraging for future applications.

5.4 Synthetic Geostationary Satellite Imagery.

Since dust is included in the RAMS-MLEF system, a new way to visualize output is needed. In Sect. 5.2, reference was made to the CRTM, which is part of GSI. Brightness temperatures (Tbs) of NWP data, void of dust, are computed by the CRTM, which are used by GSI in an assimilation process. However, since the RAMS-MLEF system contains dust, an AOD observation operator, distinct from the CRTM, was developed for the RAMS-MLEF system, which is dependent on solar reflection at 0.55 μm (see Sects. 3.2 and 5.2). A method is sought to visualize increments, which are independent of the AOD observation operator within the RAMS-MLEF system. To this end, Tbs for the SEVIRI instrument onboard MSG-08 (see Sect. 5.1) were computed, from output of the RAMS-MLEF system, with a radiative transfer model (RTM; Grasso et al. 2008), which was designed to include both moments of each of the two dust bins in the RAMS-MLEF system. Computed satellite imagery hereafter is referred to as synthetic imagery.

Several variables are needed in order to compute synthetic SEVIRI imagery. For this study, synthetic imagery was computed at both 10.80 μm and 12.00 μm, since values of Tb(10.80 μm)–Tb(12.00 μm) are useful to examine increments of simulated dust. Thus, the following two-dimensional variables were required: Latitude, longitude, and surface temperatures of both land and water bodies. Both latitude and longitude were used to compute the spectrally dependent two- dimensional surface emissivity from a monthly global dataset (Seemann et al. 2008) for the two wavelengths 10.80 and 12.00 μm. Furthermore, the following three-dimensional variables were also required: Pressure, temperature, water vapor mixing ratio along with the mass and number concentration of each dust bin. Although cloud condensate is present in RAMS-MLEF, synthetic imagery will focus exclusively on dust to avoid instances of modeled cloud layers covering and/or mixing with dust. Additional information is also needed to compute synthetic SEVIRI imagery.

In addition to modeled variables, spectrally and size dependent optical properties of dust were also required. Specifically, values of the complex index of refraction of dust, at 10.80 μm and 12.00 μm, were acquired from the Aerosol Refractive

Index Archive (ARIA; http://eodg.atm.ox.ac.uk/ARIA/, last access: 25 August 2020). Values of the complex index of refraction were used by Mie theory (Bohren and Huffman 1983) to compute the following optical properties for both wavelengths and each dust bin: Mass extinction, single-scattering albedo, and an asymmetry factor. That is, two sets of optical properties were computed; one set for *md1mp* and a second set for *md2mp*. In order for the RTM to generate synthetic MSG-08 SEVIRI imagery, the two sets of optical properties must be combined into one set, which will be referred to as the bulk set of optical properties.

Use was made of the second moment of each dust bin in order to compute the bulk set of optical properties. For example, the bulk single-scattering albedo, B_{ssa}, was computed by adding the product of the number concentration of bin 1, *md1np*, and single-scattering albedo of bin 1, ssa1, to the product of the number concentration of bin 2, *md2np*, and single-scattering albedo of bin 2, ssa2; the result was divided the sum of *md1np* + *md2np*, see Eq. (8).

$$B_{ssa} = \frac{md1np \cdot ssa1 + md2np \cdot ssa2}{md1np + md2np} \tag{8}$$

A similar number concentration weighted mean of the asymmetry factor resulted in the bulk asymmetry factor. Computation of the bulk mass extinction was slightly more involved. Values of the mass extinction coefficient for bin 1 and bin 2, from Mie, were multiplied by the mass of dust in bin 1 and bin 2, respectively to yield mass extinction. Bulk values of the mass extinction were then computed from a number concentration weighting mean of the mass extinction of each dust bin. All values of the bulk optical properties along with two- and three-dimensional variables from RAMS-MLEF were used by the RTM to generate MSG-08 SEVIRI synthetic imagery for each wavelength. Synthetic MSG-08 SEVIRI imagery at 10.80 and 12.00 µm was computed by the RTM for both the background and analysis fields. One advantage of synthetic imagery is the ability to visualize increments, which is a difference between background and analysis fields, with and without simulated dust; something that is impossible to achieve with observed imagery.

In order to evaluate model output, a comparison of simulated RAMS output with observations is necessary. Data from CALIOP (Winker et al. 2009), onboard CALIPSO, was used to produce a Vertical Feature Mask (VFM), which displays different scattering objects in the atmosphere of the Earth. For the August 2016 case herein, a descending CALIPSO ground track, white contour oriented north-northeast to south-southwest with arrows, valid about 2225 UTC 03 August 2016, is superimposed on true-color imagery from MODIS, valid near 2220 UTC 03 August 2016 (Fig. 9a). Corresponding to the CALIPSO ground track is the VFM, within which different atmospheric constituents are identified, from CALIOP (Fig. 9b). As indicated by the VFM, observed dust extended from the surface to a height of approximately 6.0 km; which is indicated by a horizontal dashed red contour. Total simulated dust mass, *md1mp* + *md2mp*, within a vertical cross section from RAMS, valid 0600 UTC 04 August 2016, green line in Fig. 9a, exhibited dust from the surface
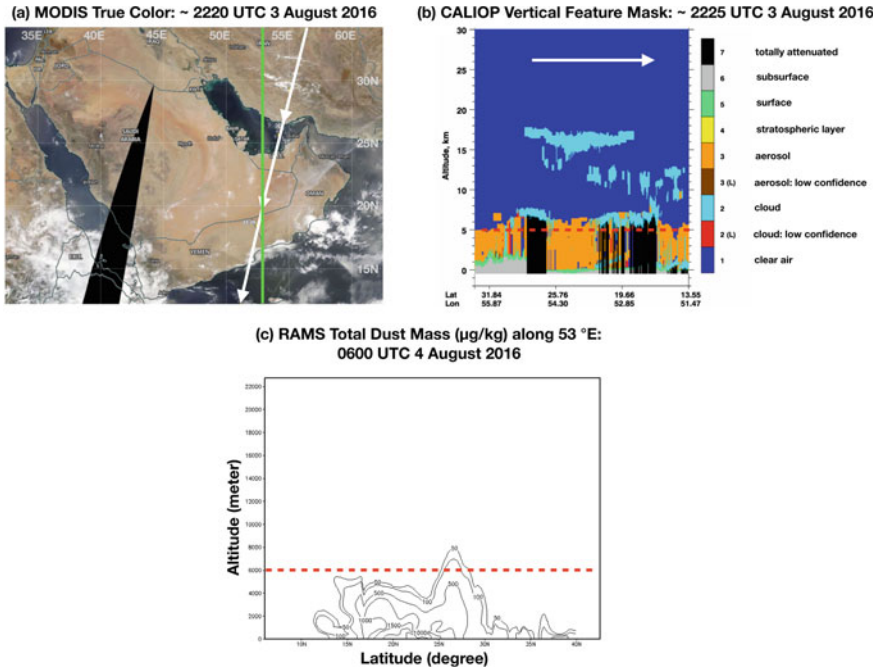
Fig. 9 a Composite true-color imagery from MODIS; the portion of the composite east of 45 E is valid at approximately 2220 UTC 3 August 2016. A white line segment with arrows denotes the ground track and motion for CALIPSO at approximately 2225 UTC 3 August 2016. A green line segment is used to denote the location of a vertical cross section from RAMS. **b** VFM from CALIOP along the CALIPSO ground track in (**a**); observed dust extended from the surface to about 6.0 km; a broken red line segment denotes a constant height of 6.0 km. **c** vertical cross section, along the green line in **a**, of the total simulated dust mass, *md1mp + md2mp,* which extended from the surface to about 6.0 km, valid at 0600 UTC 4 August 2016

to approximately 6.0 km (Fig. 9c). That is, the depth of observed dust supported the depth of dust simulated by RAMS.

Physical interpretation of increments of synthetic imagery is aided by increments of the total simulated dust mass. Dust mass of *md1mp* and *md2mp* of the background were added and then summed in the vertical throughout the depth of the simulated domain; a similar procedure was applied to the total dust mass of the analysis. Subtraction of the background dust field from the analysis dust field formed the dust increment shown in Fig. 10a. Positive (negative) regions in Fig. 10a indicated regions where dust mass was increased (decreased) as a result of the assimilation of observed AOD from MODIS. In addition to changes in *md1mp* and *md2mp* in the RAMS-MLEF assimilation system, the following three thermodynamic variables were also changed as a result of assimilation of observed AOD: Pressure, temperature, and water vapor mixing ratio (see Sect. 5.3). In order to examine the impact of increments of the three thermodynamic variables, RTM imagery at 10.80 µm (Fig. 10b) and 12.00 µm (Fig. 10c) were first produced with dust absent. Although the patterns
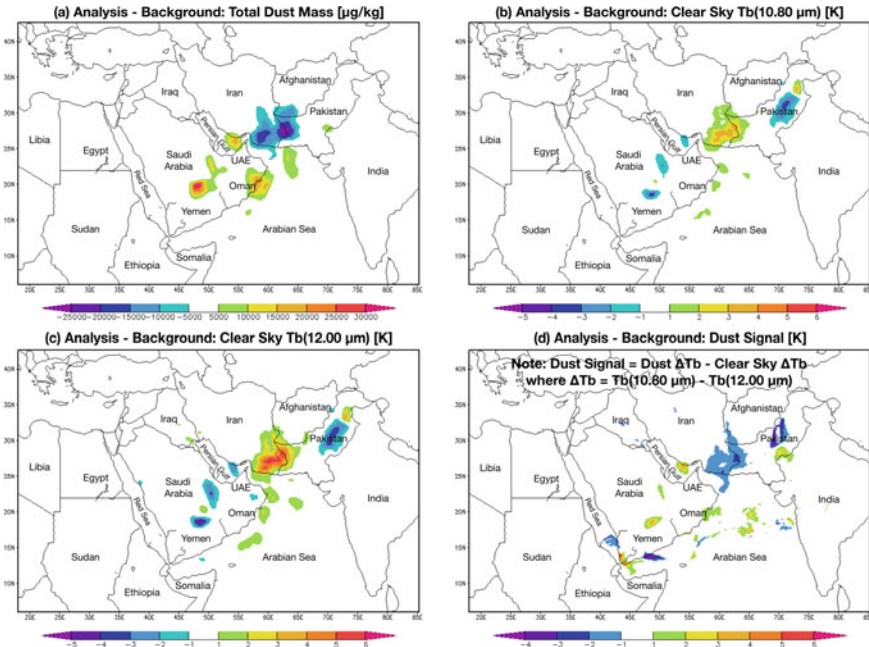
Fig. 10 Increments from RAMS-MLEF output valid at 0600 UTC 4 August 2016. Total simulated dust mass increment is shown in **a**; synthetic MGS-08 SEVIRI increments at 10.80 and 12.00 μm are displayed in **b** and **c**, respectively; increments in the dust signal are shown in **d**

evident in both Fig. 10b, c are similar, the amplitude of values were larger in synthetic imagery at 12.00 μm. Note also the opposite behavior between patterns in the total dust mass increment (Fig. 10a) and patterns in the increments of synthetic imagery at both 10.80 μm and 12.00 μm. In particular, a decrease (increase) of total dust mass resulted in an increase (decrease) of values of Tbs in imagery at both wavelengths. There were, however, regions in the synthetic increments that exhibited a lack of any relation to the dust increments; for example, central Pakistan. One possible reason for non-zero increments in synthetic imagery, that is independent of increments in dust, is a consequence of the background error covariance matrix. The background error covariance matrix spreads assimilated observations across variables and model grid points. Subsequently, increments in imagery can result as a consequence of a change in of one or more non-dust variables.

Unlike the opposite behavior between increments in total dust and increments in synthetic imagery, a similar behavior was evident between total dust increments and increments in the dust signal (Fig. 10d). In order to understand the physical interpretation of the dust signal in Fig. 10d, an explanation of how the dust signal was computed is warranted. Values of the channel difference, ΔTb = Tb(10.80 μm)–Tb(12.00 μm), may be used to detect dust; however, if the clear-sky surface is desert, then dust detection with the channel difference may be a challenge since a dust signal

may blend in with the clear-sky desert surface. One strategy, proposed herein, to isolate the dust signal is to subtract the clear-sky channel difference from the dust channel difference; that is, the dust signal is equal to $\Delta Tb_{dust} - \Delta Tb_{clear-sky}$. As a consequence, the increment in the dust signal is the dust signal of the background subtracted from the dust signal of the analysis (Fig. 10d). Regions where the dust increment in Fig. 10a increased (decreased) corresponded in an increase (decrease) in the increment of the dust signal in Fig. 10d. In particular, when the assimilation of observed AOD increased dust mass, there was a corresponding increase in the dust signal; for example, along the northern coast of the Persian Gulf, interior Sadia Arabia, border of Pakistan and India, and along the coast of Oman. In response to a reduction of total dust mass over the border of Iran and Pakistan, values of the dust signal decreased in the same region. There were also regions of values of the increment of the dust signal that showed little relationship to increments in the total dust mass. For example, there was a negative increment of the dust signal over central Pakistan, which may be a result of the background error covariance matrix. As a way to link this section with Sect. 5.3, patterns of increments of the dust signal (Fig. 10b) were similar to patterns of increments in AOD (Fig. 7d in Sect. 5.3).

5.5   *Model Response to Adjustments from Data Assimilation.*

As discussed in Sect. 5.3, the ATMAOD experiment assimilates both atmospheric and aerosol observations and updates a list of control variables as part of the analysis step of each six-hourly assimilation cycle. Other RAMS prognostic variables will respond to the changes in the control variables throughout the forecast step of the next data assimilation cycle by the model dynamical core and physical parameterizations (e.g. microphysical scheme, radiation scheme, etc.). With that, this section focuses on shedding light on the following question: What is the difference between a short-term forecast from a background initial state; that is, prior to AOD assimilation, and a forecast from an analysis initial state; that is, after AOD assimilation? In particular, this section discusses the influence of the modified total dust mass (*md1mp* + *md2mp*), which resulted from AOD assimilation, on the hydrometeor condensate field and shortwave outgoing energy. To this end, a few definitions are in order: Two simulations were conducted: (1) a simulation initialized from an analysis, which resulted from the assimilation of AOD, and is referred to as the Assimilation Forecast (AF), and (2) a simulation initialized from a background, from which an analysis is derived, and is referred to as the Background Forecast (BF). Both the AF and BF began at 0600 UTC 04 August 2016. Focus will be given to values of the Vertically Integrated Total Dust Mass (VITDM) of the BF subtracted from values of the VITDM of the AF (shaded in Fig. 11). Thus, positive values of the VITDM in Fig. 11 indicated that the assimilation of AOD increased the total dust mass in the AF compared to the BF. There are five regions in Fig. 11, within which the influence of assimilation of AOD on total condensate is discussed presently.
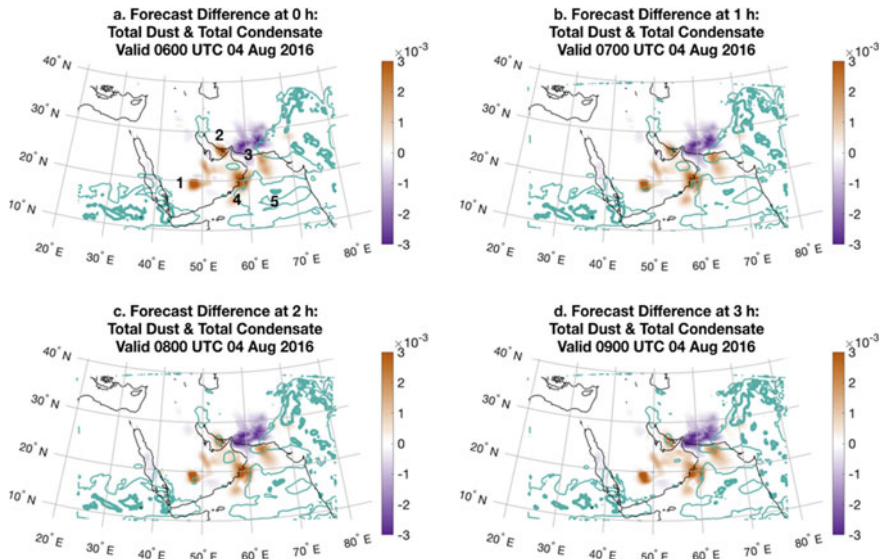
Fig. 11 Vertically integrated total dust mass ($md1mp + md2mp$; kg m^{-2}) difference between the BF and AF simulations (shaded; AF minus BF) at **a** 0 h forecast, **b** 1 h forecast, **c** 2 h forecast, and **d** 3 h forecast initialized from 0600 UTC 04 August 2016. The two green contours are used to indicate values of vertically integrated total condensate mass (mm) of a simulation initialized with an analysis field: thin for 0.1 mm; thick for 1.0 mm

Two responses of the assimilation of AOD on simulated total condensate are identified: direct and indirect. Focus will be given to regions 1, 2 (the Saudi plume and the Persian plume, respectively (see Sects. 5.1, 5.3, 5.4, and 5.5) in Fig. 11a. A plausible direct response occurred in regions 1, 2, 3, and 4 while a plausible indirect response occurred in region 5. A direct response occurred from the following: Assimilation of AOD resulted in an increase of values of the $md1mp + md2mp$ in the AF, which subsequently leads to a modification of the total number concentration, since only dust mass is a control variable (see Sect. 5.3), which then resulted in an increase of the population of Cloud Condensation Nuclei (CCN). That is, given a fixed dust particle size, an increase in dust mass, due to assimilation of AOD, will cause an increase in the dust number concentration. Development of simulated condensate occurs in RAMS when supersaturation increases above a critical value. Supersaturation is a function of upward vertical motion; therefore, when upward vertical motion occurs, supersaturation may increase above a critical value. Once supersaturations increase above a critical value, a certain percentage of the CCN population is activated to become cloud droplets, which begins a complex interaction of simulated microphysical habit types. One simulated hour after the AF simulation began, 0.1 mm of vertically integrated total condensate developed in regions 1, 2, and 4 at 0700 UTC 04 August 2016 (Fig. 11b). A progression occurred in region 3 where the 0.1 mm contour moved westward, bounding a local maximum of total

dust mass, by 0800 UTC (Fig. 11c), followed by a closed contour of 0.1 mm of vertically integrated total condensate at 0900 UTC (Fig. 11d). Notice in Fig. 11, region 5 was characterized by small changes in values of the VITDM. In response to complex changes of control variables, in region 5, through the horizontal spread of information from the flow-dependent background error covariance matrix during the assimilation of AOD at 0600 UTC 04 August 2016, temporal changes in simulated total condensate (Figs. 11 a–d), in region 5, occurred as an example of an indirect response to the assimilation of AOD. In the interest of brevity, a plausible explanation of both the direct and indirect response to the assimilation of AOD on total condensate was provided above.

Direct and indirect responses of the assimilation of AOD on simulated solar reflection are also identified. An additional consequence of increased VITDM (Fig. 11 shaded) in the AF compared to the BF was an increase in the outgoing shortwave energy (Fig. 12). That is, direct and indirect responses of the assimilation of AOD on the simulated energy budget are presently discussed. Although values of the control variable θ_{il} are prognostic, values of surface potential temperature are diagnostic. Consequently, a forecast must begin in order for the surface potential temperature to be diagnosed; thus, the time of 0610 UTC in Figs. 12 a and b. This discussion will
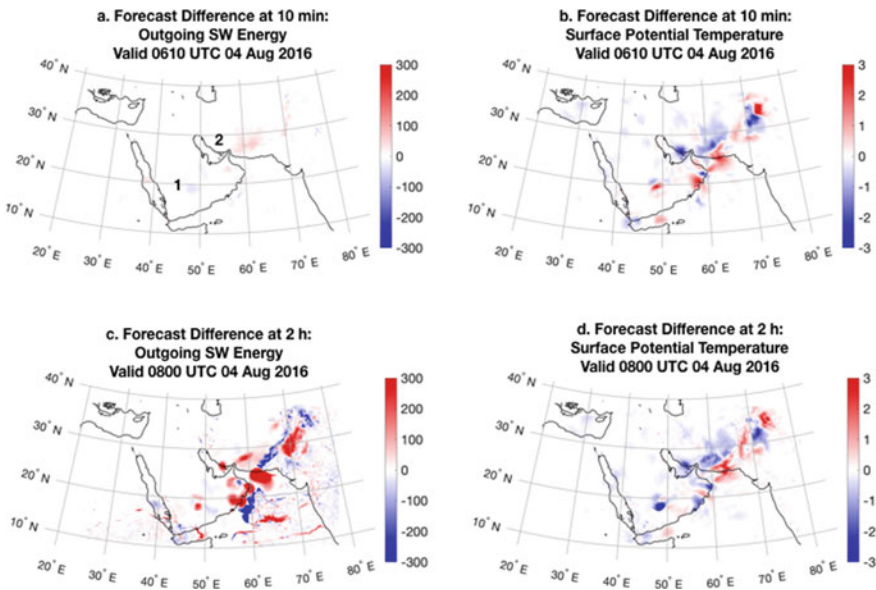

Fig. 12 a Values of the difference of outgoing simulated shortwave energy (W m^{-2}) computed from the BF subtracted from the AF at 10 min forecast initialized from 0600 UTC 4 August 2016. Positive values (red) indicate more outgoing shortwave from the AF compared to the BF simulations. **b** Values of the difference of simulated surface potential temperature (K) from the BF subtracted from the AF also at 10 min forecast initialized from 0600 UTC 4 August 2016. Negative values (blue) indicate cooler surface potential temperature from the AF compared to the BF simulations. **c–d** same as **a–b**, except for 2 h forecast

focus primarily on the Saudi plume and Persian plume, regions 1 and 2, respectively (Fig. 12a). At 0610 UTC slight variations of outgoing shortwave energy resulted from the assimilation of AOD (Fig. 12a). However, rather significant changes of surface potential temperature were already evident ten minutes into the AF simulation throughout the domain in regions away from 1 and 2 (Fig. 12b). As seen in Fig. 12b, the pattern of changes in surface potential temperature exhibited little resemblance to the pattern seen in Fig. 12a. A lack of similarity in patterns between Figs. 12a, b suggests that the influence of the flow-dependent background error covariance matrix may have been responsible for the patterns in surface potential temperature differences between the AF and BF simulations; that is, an indirect response of the assimilation of AOD on surface potential temperatures. In time, the enhanced reflection of shortwave energy from the dust mass in regions 1 and 2, evident in Fig. 12c, caused a reduction, or cooling, of the surface potential temperature at 1800 UTC. In other words, the loss of solar energy from the AF simulation, compared to the BF simulation, resulted in surface cooling below the enhanced VITDM for both regions 1 and 2 (Fig. 12 d); that is, a direct response of the assimilation of AOD on surface potential temperatures.

Although the above explanations are speculative, a more detailed analysis is, unfortunately, beyond the scope of this chapter. That is, demonstrating a link between cross-component control variables would require a thorough analysis on the role of the flow-dependent background error covariance matrix, which is responsible for updating values of control variables. That said, efforts in this section focused on providing plausible explanations for direct and indirect responses of the condensate and shortwave radiation fields to changes in total dust mass (and number concentration diagnosed afterwards) due to the assimilation of AOD.

6 Summary and Future Directions

As pointed out in Carrassi et al. (2018), coupled data assimilation is one of the major areas of active research in the field of geosciences and is expected to be advanced quickly in the coming future. In this chapter, theoretical and practical aspects of strongly coupled data assimilation with a focus on the aerosol and atmosphere coupling are discussed. We began this chapter by providing an overview and description of coupled data assimilation followed by an example from a single observation experiment of an aerosol-atmosphere coupled data assimilation using WRF-Chem. In Sect. 2, the current status of aerosol-atmosphere coupled data assimilation in both operational and research communities are reviewed in detail. Next, a description of available observational data of aerosols from various measurements such as AOD, satellite radiances, LIDAR backscattering, etc., along with a discussion of observational errors is given in Sect. 3. In Sect. 4, we present several major challenges associated with coupled data assimilation with a focus on aerosol applications. For example, the choice of control variable and the associated background error covariance is essential for the result of a successful coupled data assimilation. We

further provide a brief discussion on extending coupled data assimilation to include non-Gaussian and/or non-linear features as aerosols and their associated errors are known to behave as such. In addition, unlike meteorological observations, aerosols are under sampled. The lack of independent observations that can be used to verify the result from assimilating available aerosol observations is an issue that remains to be addressed by an improved observation network. Finally, we introduced the newly developed RAMS-MLEF, a strongly coupled aerosol-atmosphere data assimilation system, for the first time to study the impact of assimilating AOD under a strongly coupled system. A well-explored dust storm event over the Arabian Peninsula that occurred on 3–4 August 2016 was used as a case study to demonstrate the utility of the RAMS-MLEF system. In addition to examining analysis increments, which is a common practice in data assimilation, we use synthetic satellite imagery to further highlight the impact of aerosols from the viewpoint of satellite. Since short-term forecast is part of a typical data assimilation cycle, we also look into the response of aerosols adjustment from data assimilation during the short-term forecast. To end this chapter, a few future directions for research are provided.

Overall, more detailed assessments on the value of strongly coupled aerosol-atmosphere data assimilation is required. In particular, it is important for such assessments to be conducted under operational settings in order to examine more case studies with realistic configuration. In doing so, there is an urgent need to further address possibilities to improve the estimation of coupled background error covariance. While estimating coupled background error covariance under ensemble based framework may be straightforward, more work is required in order to accurately represent cross-component and cross-variable correlations for the variational aspect of hybrid based data assimilation methods (Ménard et al. 2019). In addition, using information theory to diagnose the degrees of coupling strength between any pairs of selected model variables within a coupled system can help choose control variables that are more relevant to the coupled system. Knowing the degrees of coupling strength can also benefit the efficiency of coupled data assimilation via simplifying portions of the background error covariance matrix due to low coupling strength and thus reducing computational cost. Provided that the background error covariance dictates the analysis increments, understanding the characteristics of the spatial and temporal scales of the physical processes within a coupled system is critical for assigning proper localization lengths between cross-component and cross-variable terms in the background error covariance matrix. As data assimilation methodologies advance, observations of aerosols and their corresponding observation operators also require more further development. For example, a recent study by Zhang et al. (2019) explored the use of artificial light sources for aiding AOD retrievals over nighttime. In the meantime, increasing temporal observation frequency as well as deploying instruments that allow observations of fine vertical distribution of aerosols are of critical values for improving our understanding of the spatiotemporal distribution of aerosol. There also exists a need to investigate the pros and cons of assimilation of satellite radiances sensitive to aerosols versus assimilation of retrieved quantities. Given that Artificial Intelligence (AI) techniques have shown promising results on emulating the atmosphere with sufficient training and data, there is potential to use

AI to facilitate and speed up the performance of aerosol assimilation via improved observation operators. Last but not least, verification of aerosol analysis and forecast using independent observations will benefit most from the availability of new types of observations and dense observational networks of aerosols.

References

Anderson JL (2001) An Ensemble Adjustment Kalman Filter for Data Assimilation. Mon Weather Rev 129:2884–2903

Auligné T, McNally AP, Dee DP (2007) Adaptive bias correction for satellite data in a numerical weather prediction system. Q J R Meteorol Soc 133:631–642. https://doi.org/10.1002/qj.56

Bannister RN (2008a) A review of forecast error covariance statistics in atmospheric variational data assimilation. I: characteristics and measurements of forecast error covariances. Q J R Meteorol Soc 134:1951–1970.https://doi.org/10.1002/qj.339

Bannister A (2008b) A review of forecast error covariance statistics in atmospheric variational data assimilation. II: modelling the forecast error covariance statistics. Q J R Meteorol Soc 134:1971–1996. https://doi.org/10.1002/qj.340

Benedetti A, Coauthors (2009) Aerosol analysis and forecast in the European centre for medium-range weather forecasts integrated forecast system: 2. data assimilation. J Geophys Res Atmos 114:1–18.https://doi.org/10.1029/2008JD011115

Benedetti A, Coauthors (2018) Status and future of numerical atmospheric aerosol prediction with a focus on data requirements. Atmos Chem Phys 18:10615–10643.https://doi.org/10.5194/acp-18-10615-2018

Benedetti A, Fisher M (2007) Background error statistics for aerosols. Q J R Meteorol Soc 133:391–405. https://doi.org/10.1002/qj.37

Benedetti A, Di Giuseppe F, Jones L, Peuch VH, Rémy S, Zhang X (2019) The value of satellite observations in the analysis and short-range prediction of Asian dust. Atmos Chem Phys 19:987–998.https://doi.org/10.5194/acp-19-987-2019

Bishop CH, Etherton BJ, Majumdar SJ (2001) Adaptive sampling with the ensemble transform Kalman filter Part I: theoretical aspects. Mon Weather Rev 129:420–436. https://doi.org/10.1175/1520-0493(2001)129%3c0420:ASWTET%3e2.0.CO;2

Bohren CF, Huffman DR (1983) Absorption and Scattering of Light by Small Particles. Wiley, Hoboken, p 544

Carrassi A, Bocquet M, Bertino L, Evensen G (2018) Data assimilation in the geosciences: an overview of methods, issues, and perspectives. Wiley Interdiscip Rev Clim Chang 9.https://doi.org/10.1002/wcc.535

Chen X, Andronova N, Van Leer B, Penner JE, Boyd JP, Jablonowski C, Lin SJ (2013) A control-volume model of the compressible Euler equations with a vertical lagrangian coordinate. Mon Weather Rev 141:2526–2544.https://doi.org/10.1175/MWR-D-12-00129.1

Chikhar K, Gauthier P (2017) Impact of lateral boundary conditions on regional analyses. Mon Weather Rev 145:1361–1379. https://doi.org/10.1175/MWR-D-16-0245.1

Chin M, Rood RB, Lin SJ, Müller JF, Thompson AM (2000) Atmospheric sulfur cycle simulated in the global model GOCART: model description and global properties. J Geophys Res Atmos 105:24671–24687. https://doi.org/10.1029/2000JD900384

Choi Y, Chen SH, Huang CC, Earl K, Chen CY, Schwartz CS, Matsui T (2020) Evaluating the impact of assimilating aerosol optical depth observations on dust forecasts over North Africa and the East Atlantic using different data assimilation methods. J Adv Model Earth Syst 12:1–30. https://doi.org/10.1029/2019MS001890

Collins WJ, Coauthors (2011) Development and evaluation of an earth-system model—HadGEM2. Geosci Model Dev 4:1051–1075.https://doi.org/10.5194/gmd-4-1051-2011

Collins WD, Rasch PJ, Eaton BE, Khattatov BV, Lamarque JF, Zender CS (2001) Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: methodology for INDOEX. J Geophys Res Atmos 106:7313–7336. https://doi.org/10.1029/2000JD 900507

Cotton WR, Coauthors (2003) RAMS 2001: current status and future directions. Meteorol Atmos Phys 82:5–29.https://doi.org/10.1007/s00703-001-0584-9

Dee DP (2005) Bias and data assimilation. Q J R Meteorol Soc 131:3323–3343. https://doi.org/10.1256/qj.05.137

Derber JC, Wu W-S (1998) The Use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. Mon Weather Rev 126:2287–2299. https://doi.org/10.1175/1520-0493(1998)126%3c2 287:TUOTCC%3e2.0.CO;2

Derber JC, Wu W-S, Parrish DF, Lord SJ (1991) The new global operational analysis system at the national meteorological center. Wea Forecast 6:538–547

Descombes G, Auligné T, Vandenberghe F, Barker DM, Barré J (2015) Generalized background error covariance matrix model (GEN-BE v2.0). Geosci Model Dev 8:669–696. https://doi.org/10.5194/gmd-8-669-2015

Eltahan M, Alahmadi S (2019) Numerical dust storm simulation using modified geographical domain and data Assimilation: 3DVAR and 4DVAR (WRF-Chem/WRFDA). IEEE Access 7:128980–128989. https://doi.org/10.1109/ACCESS.2019.2930812

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. J Geophys Res 99:10143–10162

Eyre JR (1992) A bias correction scheme for simulated TOVS brightness temperatures

Fast JD, Gustafson WI, Easter RC, Zaveri RA, Barnard JC, Chapman EG, Grell GA, Peckham SE (2006) Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model. J Geophys Res Atmos 111:1–29. https://doi.org/10.1029/2005JD006721

Federico S, Petracca M, Panegrossi G, Dietrich S (2017) Improvement of RAMS precipitation forecast at the short-range through lightning data assimilation. Nat Hazards Earth Syst Sci 17:61–76. https://doi.org/10.5194/nhess-17-61-2017

Fisher M (2003) Background error covariance modelling. Semin Recent Dev Data Assim 45–63

Fisher M (2006) Wavelet Jb—a new way to model the statistics of background errors. ECMWF Newsl Winter 200:23–28. https://doi.org/10.21957/ef5601ak43

Fletcher SJ, Jones AS (2014) Multiplicative and additive incremental variational data assimilation for mixed lognormal-Gaussian errors. Mon Weather Rev 142:2521–2544. https://doi.org/10.1175/MWR-D-13-00136.1

Grasso LD, Sengupta M, Dostalek JF, Brummer R, Demaria M (2008) Synthetic satellite imagery for current and future environmental satellites. Int J Remote Sens 29:4373–4384. https://doi.org/10.1080/01431160801891820

Grell G, Baklanov A (2011) Integrated modeling for forecasting weather and air quality: a call for fully coupled approaches. Atmos Environ 45:6845–6851. https://doi.org/10.1016/j.atmosenv.2011.01.017

Grell GA, Peckham SE, Schmitz R, McKeen SA, Frost G, Skamarock WC, Eder B (2005) Fully coupled online chemistry within the WRF model. Atmos Environ 39:6957–6975. https://doi.org/10.1016/j.atmosenv.2005.04.027

Guth J, Josse B, Marécal V, Joly M, Hamer P (2016) First implementation of secondary inorganic aerosols in the MOCAGE version R2.15.0 chemistry transport model. Geosci Model Dev 9:137–160. https://doi.org/10.5194/gmd-9-137-2016

Hakami A, Henze DK, Seinfeld JH, Chai T, Tang Y, Carmichael GR, Sandu A (2005) Adjoint inverse modeling of black carbon during the Asian Pacific Regional Aerosol Characterization Experiment. J Geophys Res D Atmos 110:1–17. https://doi.org/10.1029/2004JD005671

Han Y, van Delst P, Liu Q, Weng F, Yan B, Treadon R, Derber J (2006) NOAA technical report NESDIS 122 JCSDA Community Radiative Transfer Model (CRTM)—Version 1. NOAA Tech Rep

Hogan TF, Coauthors (2014) The navy global environmental model. Ocenography 27:116–125.https://doi.org/10.1017/CBO9781107415324.004

Houtekamer PL, Mitchell HL (2001) A sequential ensemble Kalman filter for atmospheric data assimilation. Mon Weather Rev 129:123–137. https://doi.org/10.1175/1520-0493(2001)129%3c0123:ASEKFF%3e2.0.CO;2

Hsu NC, Tsay SC, King MD, Herman JR (2006) Deep Blue retrievals of Asian aerosol properties during ACE-Asia. IEEE Trans Geosci Remote Sens 44:3180–3195. https://doi.org/10.1109/TGRS.2006.879540

Hu M, Xue M (2007) Impact of configurations of rapid intermittent assimilation of WSR-88D radar data for the 8 May 2003 Oklahoma City tornadic thunderstorm case. Mon Weather Rev 135:507–525. https://doi.org/10.1175/MWR3313.1

Jish Prakash P, Stenchikov G, Kalenderski S, Osipov S, Bangalath H (2015) The impact of dust storms on the Arabian Peninsula and the Red Sea. Atmos Chem Phys 15:199–222. https://doi.org/10.5194/acp-15-199-2015

Kahnert M (2008) Variational data analysis of aerosol species in a regional CTM: background error covariance constraint and aerosol optical observation operators. Tellus B 60:753–770

Kalnay E (2003) Atmospheric modeling data assimilation and predictability. Cambridge University Press, Cambridge, p 341

Kleist DT, Ide K (2015) An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. Mon Weather Rev 143:452–470. https://doi.org/10.1175/MWR-D-13-00350.1

Kleist DT, Ide K, Parrish DF, Derber JC, Treadon R, Wu W-S, Lord S (2009) Introduction of the GSI into the NCEP Global Data Assimilation System. Weather Forecast 24:1691–1705.https://doi.org/10.1175/2009WAF2222201.1

Kliewer AJ, Fletcher SJ, Jones AS, Forsythe JM (2016) Comparison of gaussian, logarithmic transform and mixed Gaussian-log-normal distribution based 1DVAR microwave temperature-water-vapour mixing ratio retrievals. Q J R Meteorol Soc 142:274–286. https://doi.org/10.1002/qj.2651

Lee E, Županski M, Županski D, Park SK (2017) Impact of the OMI aerosol optical depth on analysis increments through coupled meteorology–aerosol data assimilation for an Asian dust storm. Remote Sens Environ 193:38–53. https://doi.org/10.1016/j.rse.2017.02.013

Levy RC, Mattoo S, Munchak LA, Remer LA, Sayer AM, Patadia F, Hsu NC (2013) The collection 6 MODIS aerosol products over land and ocean. Atmos Meas Tech 6:2989–3034. https://doi.org/10.5194/amt-6-2989-2013

Liu Z, Liu Q, Lin HC, Schwartz CS, Lee YH, Wang T (2011) Three-dimensional variational assimilation of MODIS aerosol optical depth: Implementation and application to a dust storm over East Asia. J Geophys Res Atmos 116:1–19. https://doi.org/10.1029/2011JD016159

Lorenc AC (1986) Analysis methods for numerical weather prediction. Q J R Meteorol Soc 112:1177–1194. https://doi.org/10.1002/qj.49711247414

Lynch P, Coauthors (2016) An 11-year global gridded aerosol optical thickness reanalysis (v1.0) for atmospheric and climate sciences. Geosci Model Dev 9:1489–1522.https://doi.org/10.5194/gmd-9-1489-2016

Ménard R, Gauthier P, Rochon Y, Robichaud A, de Grandpré J, Yang Y, Charrette C, Chabrillat S (2019) Coupled stratospheric chemistry-meteorology data assimilation. Part II: weak and strong coupling. Atmosphere (Basel) 10. https://doi.org/10.3390/ATMOS10120798

Miller SD, Coauthors 2019 A tale of two dust storms: analysis of a complex dust event in the Middle East. Atmos Meas Tech 12:5101–5118.https://doi.org/10.5194/amt-12-5101-2019

Molod A, Takacs L, Suarez M, Bacmeister J, Song I-S, Eichmann A (2012) The GEOS-5 atmospheric general circulation model: mean climate and development from MERRA to fortuna

Morcrette JJ, Beljaars A, Benedetti A, Jones L, Boucher O (2008) Sea-salt and dust aerosols in the ECMWF IFS model. Geophys Res Lett 35:1–5. https://doi.org/10.1029/2008GL036041

Morcrette JJ, Beljaars A, Benedetti A, Jones L, Boucher O, Coauthors (2009) Aerosol analysis and forecast in the european centre for medium-range weather forecasts integrated forecast system: Forward modeling. J Geophys Res Atmos 114:1–17. https://doi.org/10.1029/2008JD011235

Pagowski M, Grell GA (2012) Experiments with the assimilation of fine aerosols using an ensemble Kalman filter. J Geophys Res Atmos 117:1–15. https://doi.org/10.1029/2012JD018333

Pagowski M, Liu Z, Grell GA, Hu M, Lin HC, Schwartz CS (2014) Implementation of aerosol assimilation in gridpoint statistical interpolation (v. 3.2) and WRF-Chem (v. 3.4.1). Geosci Model Dev 7:1621–1627. https://doi.org/10.5194/gmd-7-1621-2014

Parrish DF, Derber JC (1992) The national meteorological center's spectral statistical-interpolation analysis system. Mon Weather Rev 120:1747–1763. https://doi.org/10.1175/1520-0493(1992)120%3c1747:TNMCSS%3e2.0.CO;2

Penny SG, Hamill TM (2017) Coupled data assimilation for integrated earth system analysis and prediction. Bull Am Meteorol Soc 98:169–172. https://doi.org/10.1175/BAMS-D-17-0036.1

Penny SG, Hamill TM, Coauthors (2017) Coupled data assimilation for integrated earth system analysis and prediction: goals, challenges and recommendations, p 50. www.wmo.int/pages/prog/arep/wwrp/new/documents/Final_WWRP_2017_3_27_July.pdf

Petters MD, Kreidenweis SM (2007) A single parameter representation of hygroscopic growth and cloud condensation nucleus activity-Part 3: including surfactant partitioning. Atmos Chem Phys 7:1961–1971. https://doi.org/10.5194/acp-13-1081-2013

Putman WM, Lin SJ (2007) Finite-volume transport on various cubed-sphere grids. J Comput Phys 227:55–78. https://doi.org/10.1016/j.jcp.2007.07.022

Rabier F, Jarvinen H, Klinker E, Mahfouf J-F, Simmons A (1999) The ECMWF operational implementationof four dimensional variational assimilation. Part I: experimental results with simplified physics

Randles CA, Coauthors (2017) The MERRA-2 aerosol reanalysis, 1980 onward. Part I: description and data assimilation evaluation. J Clim 30: 6823–6850. https://doi.org/10.1175/JCLI-D-16-0609.1

Remer LA, Coauthors (2005) The MODIS aerosol algorithm, products, and validation. J Atmos Sci 62:947–973.https://doi.org/10.1175/JAS3385.1

Rubin JI, Coauthors (2016) Development of the Ensemble Navy Aerosol Analysis Prediction System (ENAAPS) and its application of the Data Assimilation Research Testbed (DART) in support of aerosol forecasting. Atmos Chem Phys 16:3927–3951.https://doi.org/10.5194/acp-16-3927-2016

Rubin JI, Reid JS, Hansen JA, Anderson JL, Holben BN, Xian P, Westphal DL, Zhang J (2017) Assimilation of AERONET and MODIS AOT observations using variational and ensemble data assimilation methods and its impact on aerosol forecasting skill. J Geophys Res 122: 4967–4992.https://doi.org/10.1002/2016JD026067

Saleeby SM, Cotton WR (2004) A large-droplet mode and prognostic number concentration of cloud droplets in the colorado state university regional atmospheric modeling system (RAMS). Part I: module descriptions and supercell test simulations. J Appl Meteorol 43:182–195. https://doi.org/10.1175/1520-0450(2004)043%3c0182:ALMAPN%3e2.0.CO;2

Saleeby SM, van den Heever SC (2013) Developments in the CSU-RAMS aerosol model: emissions, nucleation, regeneration, deposition, and radiation. J Appl Meteorol Climatol 52:2601–2622.https://doi.org/10.1175/JAMC-D-12-0312.1

Saleeby SM, Cotton WR, van den Heever SC, Coauthors (2019) The influence of simulated surface dust lofting erodible fraction on radiative forcing. Atmos Chem Phys 19:10279–10301.https://doi.org/10.5194/acp-19-10279-2019

Sandu A, Liao W, Carmichael GR, Henze DK, Seinfeld JH (2005) Inverse modeling of aerosol dynamics using adjoints: theoretical and numerical considerations. Aerosol Sci Technol 39:677–694. https://doi.org/10.1080/02786820500182289

Schwartz CS, Liu Z, Lin H-C, Cetola JD (2014) Assimilating aerosol observations with a "hybrid" variational-ensemble data assimilation system. J Geophys Res Atmos 119:4043–4069.https://doi.org/10.1002/2013JD020937.Received

Seemann SW, Borbas EE, Knuteson RO, Stephenson GR, Huang HL (2008) Development of a global infrared land surface emissivity database for application to clear sky sounding retrievals from multispectral satellite radiance measurements. J Appl Meteorol Climatol 47:108–123. https://doi.org/10.1175/2007JAMC1590.1

Sekiyama TT, Tanaka TY, Shimizu A, Miyoshi T (2010) Data assimilation of CALIPSO aerosol observations. Atmos Chem Phys 10:39–49. https://doi.org/10.5194/acp-10-39-2010

Sun J (2006) Convective-scale assimilation of radar data: Progress and challenges. Q J R Meteorol Soc 131:3439–3463. https://doi.org/10.1256/qj.05.149

Sun J, Wang H, Tong W, Zhang Y, Lin C-Y, Xu D (2015) Comparison of the Impacts of momentum control variables on high-resolution variational data assimilation and precipitation forecasting. Mon Weather Rev 144:149–169.https://doi.org/10.1175/MWR-D-14-00205.1

Suzuki K, Zupanski M (2018) Uncertainty in solid precipitation and snow depth prediction for Siberia using the Noah and Noah-MP land surface models. Front Earth Sci 12:672–682

Tanaka TY, Chiba M (2005) Global simulation of dust aerosol with a chemical transport model MASINGAR. J Meteorol Soc Japan 83:255–278. https://doi.org/10.2151/jmsj.83a.255

Tang Y, Coauthors (2009) The impact of chemical lateral boundary conditions on CMAQ predictions of tropospheric ozone over the continental United States Environ Fluid Mech 9:43–58.https://doi.org/10.1007/s10652-008-9092-5

Thépaut J-N, Courtier P, Belaud G, Lemaître G (1996) Dynamical structure functions in a four-dimensional variational assimilation: a case study. Q J R Meteorol Soc 122:535–561. https://doi.org/10.1002/qj.49712253012

Leeuwen Van PJ (2010) Nonlinear data assimilation in geosciences: an extremely efficient particle filter. Q J R Meteorol Soc 136:1991–1999. https://doi.org/10.1002/qj.699

Wang H, Niu T (2013) Sensitivity studies of aerosol data assimilation and direct radiative feedbacks in modeling dust aerosols. Atmos Environ 64:208–218. https://doi.org/10.1016/j.atmosenv.2012.09.066

Weaver C, Coauthors (2007) Direct insertion of MODIS radiances in a global aerosol transport model. J Atmos Sci 64 808–827.https://doi.org/10.1175/JAS3838.1

Whitaker JS, Hamill TM (2002) Ensemble data assimilation without perturbed observations. Mon Weather Rev 1913–1924

Whitaker JS, Hamill TM (2012) Evaluating methods to account for system errors in ensemble data assimilation. Mon Weather Rev 140:3078–3089.https://doi.org/10.1175/MWR-D-11-00276.1

Winker DM, Vaughan MA, Omar A, Hu Y, Powell KA, Liu Z, Hunt WH, Young SA (2009) Overview of the CALIPSO mission and CALIOP data processing algorithms. J Atmos Ocean Technol 26:2310–2323. https://doi.org/10.1175/2009JTECHA1281.1

Wu W-S, Purser RJ, Parrish DF (2002) Three-dimensional variational analysis with spatially inhomogeneous covariances. Mon Weather Rev 130:2905–2916. https://doi.org/10.1175/1520-0493(2002)130%3c2905:TDVAWS%3e2.0.CO;2

Xian P, Coauthors (2019) Current State of the global operational aerosol multi-model ensemble: an update from the International Cooperative for Aerosol Prediction (ICAP) Q J R Meteorol Soc 176–209https://doi.org/10.1002/qj.3497

Xie Y, MacDonald AE (2012) Selection of momentum variables for a three-dimensional variational analysis. Pure Appl Geophys 169:335–351. https://doi.org/10.1007/s00024-011-0374-3

Yumimoto K, Tanaka TY, Yoshida M, Kikuchi M, Nagao TM, Murakami H, Maki T (2018) Assimilation and forecasting experiment for heavy Siberian wildfire smoke in May 2016 with himawari-8 aerosol optical thickness. J Meteorol. Soc Japan 96B:133. https://doi.org/10.2151/jmsj.2018-035

Zhang F, Snyder C Sun J (2004) Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble kalman filter. Mon Weather Rev 1238–1253

Zhang J, Campbell JR, Hyer EJ, Reid JS, Westphal DL, Johnson RS (2014) Evaluating the impact of multisensor data assimilation on a global aerosol particle transport model. J Geophys Res Atmos 119:4674–4689. https://doi.org/10.1002/2013JD020975.Received

Zhang J, Campbell JR, Hyer EJ, Reid JS, Westphal DL, Johnson RS, Jaker SL, Reid JS, Miller SD, Solbrig J, Toth TD (2019) Characterization and application of artificial light sources for nighttime

aerosol optical depth retrievals using the visible infrared imager radiometer suite day/night band. Atmos Mea 12:3209–3222

Zupanski D (1993) The effects of discontinuities in the Betts–Miller cumulus convection scheme on four-dimensional variational data assimilation. Tellus A. https://doi.org/10.3402/tellusa.v45i5.15053

Zupanski M (2005) Maximum likelihood ensemble filter: theoretical aspects. Mon Weather Rev 133:1710–1726. https://doi.org/10.1175/MWR2946.1

Zupanski M (2017) Data assimilation for coupled modeling systems. In: Park S-K, Xu L (eds) Data assimilation for atmospheric, oceanic and hydrologic applications, vol 3. Springer, Berlin, Heidelberg, pp 1–553

Zupanski M, Fletcher SJ, Navon IM, Uzunoglu B, Heikes RP, Randall DA, Ringler TD, Daescu D (2006) Initiation of ensemble data assimilation. Tellus Ser A Dyn Meteorol Oceanogr 58:159–170. https://doi.org/10.1007/s00024-011-0374-3

Zupanski M, Navon IM, Zupanski D (2008) The maximum likelihood ensemble filter as a non-differentiable minimization algorithm. Q J R Meteorol Soc 134:1039–1050